

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number
WO 01/92523 A2

(51) International Patent Classification⁷: C12N 15/12,
9/00, C12Q 1/68, C07K 14/47, 14/705, 16/18, A61K
31/7088, 38/17, 39/395, G01N 33/53

CT 06516 (US). LEACH, Martin, D. [US/US]; 131
Beechwood Drive, Madison, CT 06433 (US).

(21) International Application Number: PCT/US01/10836

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris,
Glovsky and Popeo, P.C., One Financial Center, Boston,
MA 02111 (US).

(22) International Filing Date: 29 May 2001 (29.05.2001)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(30) Priority Data:
60/206,132 30 May 2000 (30.05.2000) US
60/228,716 29 August 2000 (29.08.2000) US

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/203,132 (CIP)
Filed on 30 May 2000 (30.05.2000)
US 60/228,716 (CIP)
Filed on 29 August 2000 (29.08.2000)

(71) Applicant (for all designated States except US): CURA-GEN CORPORATION [US/US]; 11th Floor, 555 Long Wharf Drive, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): SHIMKETS, Richard, A. [US/US]; 191 Leete Street, West Haven,



WO 01/92523 A2

(54) Title: NOVEL HUMAN POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

(57) Abstract: The present invention provides ORFX, a novel isolated polypeptide, as well as a polynucleotide encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptide, polynucleotide or antibody. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.

NOVEL HUMAN POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

TECHNICAL FIELD OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom.

5

BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as “ORFX”.

The invention provides isolated nucleic acid molecules that encode novel polypeptides, with the nucleic acids and polypeptides being denoted SEQ ID NO. 1-22982. In one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an ORFX polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

- In a another aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample. The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition.
- 5 Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.
- 10

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

15

In a still further aspect, the invention provides a method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

20

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned

30

herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting. Other features and advantages of the invention will be apparent from the following detailed description and claims.

5

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are novel nucleic acid sequences and their encoded polypeptides listed in the Sequence Listing as SEQ ID NO. 1-22982. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptides are referred to as "ORFX polypeptides" or ORFX proteins". The ORFX polynucleotides and the encoded polypeptides are characterized by having novel sequences that were discovered as a result of SeqCalling™ analysis conducted on human tissues from a broad range of sources. SeqCalling™ is disclosed in U. S. Ser. No. 09/417,386, filed Oct. 13, 1999, incorporated herein by reference in its entirety. Sample preparation for SeqCalling™ can be performed by the sample preparation method described in U. S. Patent No. 5,871,697 and in Shimkets et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999), incorporated herein by reference in their entireties. In many cases the sequences disclosed herein were assembled using additional SeqCalling™ fragments.

In the designation "ORFX", the "X" can take on any of the values from 1 to 11491. For example, an ORFX nucleic acid according to the invention is a nucleic acid including a sequence such as an ORF1 nucleic acid, and an ORFX polypeptide according to the invention is a polypeptide that includes the amino acid sequence of a polypeptide such as an ORF1 polypeptide. Unless indicated otherwise, "ORFX" is meant to refer to any one, several, or all of the sequences disclosed herein. The sequences of the nucleic acids and the encoded polypeptides of the invention are disclosed in the appended Sequence Listing. More specifically, examples of the

sequences of the nucleic acids of the invention are disclosed in the appended Sequence Listing in SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743; and examples of the sequences of the polypeptides of the invention are disclosed in the appended Sequence Listing in SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5 5000, 5076, 5080, 19474, 19476, 19478, 11740, 11742 and 11744.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides.

Table 1 has six columns whose headings are as follows.

Column 1 of Table 1, entitled "Entry" denotes the ORFX number assigned to a nucleic acid and polypeptide according to the invention.

10 Column 2 of Table 1, entitled "Sequence Id", provides a CuraGen Corporation's internal identification number for the indicated ORF, along with the SEQ ID Nos (enclosed in parentheses) corresponding to the indicated ORF and the polypeptide encoded by it. For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO: 5897, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO: 5898.

15 Column 3 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. Information about the previously described proteins can be retrieved from <http://www.ncbi.nlm.nih.gov/>.

To determine similarity to previously described proteins, polypeptides encoded by ORFX 20 DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected. Open reading frames were extended beyond the region of the protein matched using standard 25 DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 4 of Table 1, entitled "Protein Domain", lists previously described protein domains, designated by pfam entries, that are present in the polypeptides encoded by the ORFs. Also included in column 3 are proteins in which these domains are present. The pfam entries can 30 be retrieved from <http://pfam.wustl.edu/>. DNA sequences were translated in all six frames and

tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at <http://pfam.wustl.edu>). Translated DNA sequences that matched a protein domain entry in the Pfam database and had a score of '7.5' were selected.

Column 5 of Table 3, entitled "Protein Classification", lists the classification assigned for 5 the protein, based on its homology. Examples of proteins in the classification include the following:

Amylases

Amylase is responsible for endohydrolysis of 1,4-alpha-glucosidic linkages in 10 oligosaccharides and polysaccharides. Variations in amylase gene may be indicative of delayed maturation and of various amylase producing neoplasms and carcinomas.

Amyloid

The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL). The synthesis of certain members 15 of the family is greatly increased in inflammation. Prolonged elevation of plasma SAA levels, as in chronic inflammation, results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. Amyloid selectively inhibits insulin-stimulated glucose utilization and glycogen deposition in muscle, while not affecting adipocyte glucose metabolism. Deposition of fibrillar amyloid proteins intraneuronally, as neurofibrillary tangles, extracellularly, as plaques 20 and in blood vessels, is characteristic of both Alzheimer's disease and aged Down's syndrome. Amyloid deposition is also associated with type II diabetes mellitus.

Angiopoietin

Members of the angiopoietin/fibrinogen family have been shown to stimulate the generation of new blood vessels, inhibit the generation of new blood vessels, and perform several 25 roles in blood clotting. This generation of new blood vessels, called angiogenesis, is also an essential step in tumor growth in order for the tumor to get the blood supply it needs to expand. Variation in these genes may be predictive of any form of heart disease, numerous blood clotting

disorders, stroke, hypertension and predisposition to tumor formation and metastasis. In particular, these variants may be predictive of the response to various antihypertensive drugs and chemotherapeutic and anti-tumor agents.

Apoptosis-related proteins

5 Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes preventing their target-cells from dying too soon. Variants of apoptosis related genes may be
10 useful in formulation of anti-aging drugs.

Cadherin, Cyclin, Polymerase, Oncogenes, Histones, Kinases

Members of the cell division/cell cycle pathways such as cyclins, many transcription factors and kinases, DNA polymerases, histones, helicases and other oncogenes play a critical role in carcinogenesis where the uncontrolled proliferation of cells leads to tumor formation and
15 eventually metastasis. Variation in these genes may be predictive of predisposition to any form of cancer, from increased risk of tumor formation to increased rate of metastasis. In particular, these variants may be predictive of the response to various chemotherapeutic and anti-tumor agents.

Colony-stimulating factor-related proteins

Granulocyte/macrophage colony-stimulating factors are cytokines that act in
20 hematopoiesis by controlling the production, differentiation, and function of 2 related white cell populations of the blood, the granulocytes and the monocytes-macrophages.

Complement-related proteins

Complement proteins are immune associated cytotoxic agents, acting in a chain reaction to exterminate target cells to that were opsonized (primed) with antibodies, by forming a membrane
25 attack complex (MAC). The mechanism of killing is by opening pores in the target cell membrane. Variations in 20 complement genes or their inhibitors are associated with many

autoimmune disorders. Modified serum levels of complement products cause edemas of various tissues, lupus (SLE), vasculitis, glomerulonephritis, renal failure, hemolytic anemia, thrombocytopenia, and arthritis. They interfere with mechanisms of ADCC (antibody dependent cell cytotoxicity), severely impair immune competence and reduce phagocytic ability. Variants of 5 complement genes may also be indicative of type I diabetes mellitus, meningitis neurological disorders such as nemaline myopathy, neonatal hypotonia, muscular disorders such as congenital myopathy and other diseases.

Cytochrome

The respiratory chain is a key biochemical pathway which is essential to all aerobic cells. 10 There are five different cytochromes involved in the chain. These are heme bound proteins which serve as electron carriers. Modifications in these genes may be predictive of ataxia areflexia, dementia and myopathic and neuropathic changes in muscles. Also, association with various types of solid tumors.

Kinesins

15 Kinesins are tubulin molecular motors that function to transport organelles within cells and to move chromosomes along microtubules during cell division. Modifications of these genes may be indicative of neurological disorders such as Pick disease of the brain, tuberous sclerosis.

Cytokines, Interferon, Interleukin

Members of the cytokine families are known for their potent ability to stimulate cell 20 growth and division even at low concentrations. Cytokines such as erythropoietin are cell-specific in their growth stimulation; erythropoietin is useful for the stimulation of the proliferation of erythroblasts. Variants in cytokines may be predictive for a wide variety of diseases, including cancer predisposition.

G-protein coupled receptors

25 G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by

interaction with guanine nucleotide-binding (G) proteins. Alterations in genes coding for G-coupled proteins may be involved in and indicative of a vast number of physiological conditions. These include blood pressure regulation, renal dysfunctions, male infertility, dopamine associated cognitive, emotional, and endocrine functions, hypercalcemia, chondrodysplasia and osteoporosis, 5 pseudohypoparathyroidism, growth retardation and dwarfism.

Thioesterases

Eukaryotic thiol proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Variants of thioester 10 associated genes may be predictive of neuronal disorders and mental illnesses such as Ceroid Lipofuscinosis, Neuronal 1, Infantile, Santavuori disease and more.

The key to the molecule type is as follows:

15	Abbrev:	Title:
	amylase	amylase protein
	amylaseinhib	amylase inhibitor
	amyloid	amyloid protein
20	apoptosis	apoptosis associated protein
	apoptosisinhib	apoptosis inhibitors
	apoptosisrecep	apoptosis receptors
	ATPase_associated	ATPase associated protein
	biotindep	biotin dependent enzyme/protein
25	cadherin	cadherin protein
	calcium_channel	calcium channel protein
	carboxylase	carboxylase protein
	cathepsin	cathepsin/carboxypeptidases
	cathepsininhib	cathepsin/carboxypeptidase inhibitor
30	chloride_channel	chloride channel protein
	collagen	collagen
	complement	complement protein
	complementrecept	complement receptor protein
	complementinhib	complement inhibitor
35	csf	colony stimulating factor
	csfrecept	colony stimulating factor receptor

	cyclin	cyclin protein
	cyto450	cytochrome p450 protein
	cytochrome	cytochrome related protein
	deaminase	deaminase
5	dehydrogenase	dehydrogenase
	desaturase	desaturase
	dna_ma_bind	DNA/RNA binding protein/factor
	dna_ma_inhib	DNA/RNA binding protein/factor inhibitor
	dynein	dynein
10	elastase	elastase
	elastaseinhib	elastase inhibitor
	eph	EPH family of tyrosine kinases
	esterase	esterase
	esteraseinhib	esterase inhibitor
15	fgf	fibroblast growth factor
	fgfreceptor	fibroblast growth factor receptor
	gaba	GABA receptor
	glucoamylase	glucoamylase
	glucoronidase	glucuronidase
20	glycoprotein	glycoprotein
	Guanylyl	guanylylate cyclase
	helicase	helicase
	histone	histone
	HOM	homologous
25	homeobox	homeobox protein
	hydrolase	hydrolase
	hydroxysteroid	hydroxysteroid associated protein
	hypoxanthine	hypoxanthine associated protein
	immunoglob	immunoglobulin
30	immunoglobrecept	immunoglobulin receptor
	interferon	interferon
	interleukin	interleukin
	interleukinrecept	interleukin receptor
	isomerase	isomerase
35	isomeraseinhibitor	isomerase inhibitor
	isomeraseseceptor	isomerase receptor
	kinase	kinase
	kinaseinhibitor	kinase inhibitor
	kinasereceptor	kinase receptor
40	kinesin	kinesin
	laminin	laminin associated protein
	lipase	lipase
	metallothionein	metallothionein
	MHC	major histocompatibility complex
45	misc_channel	miscellaneous channel

	ngf	nerve growth factor
	nuci_recpt	nuclear receptor
	nuclease	nuclease
	oncogene	oncogene associated protein
5	oxidase	oxidase
	oxygenase	oxygenase
	peptidase	peptidase
	peroxidase	peroxidase
	phosphatase	phosphatase
10	phosphataseinhib	phosphatase inhibitor
	phosphorylase	phosphorylase
	PIR	PIR DATABASE (release 56, 29-OCT-1998)
	polymerase	polymerase
	potassium_channel	potassium channel protein
15	prostaglandin	prostaglandin
	protease	protease
	proteaseinhib	protease inhibitor
	reductase	reductase
	ribosomalprot	ribosomal associated protein
20	RTR	EMBLDATABASE translated entries not to be incorporated into SWISS-PROT (20-JUL-1998)
	SIM	similar
	SPTR	EMBL DATABASE translated entries to be incorporated into SWISS-PROT (20-JUL-1998)
25	struct	structural associated protein
	sulfotransferase	sulfotransferase
	SWP	SWISS-PROT DATABASE (release 18-OCT-1998)
30	SWPN	SWISS-PROT Update (release 11-NOV-98)
	synthase	synthase
	tgf	transforming growth factor
	tgfreceptor	transforming growth factor receptor
35	thioesterase	thioesterase
	thiolase	thiolase
	tm7	seven transmembrane domain G-protein coupled receptor
	tnf	necrosis factor receptor
40	traffic	tumor necrosis factor
	tnfreceptor	tumor trafficking associated protein
	TRN	EMBL DATABASE translated entries update (20-JUL-1998)
	transcriptfactor	transcription factor
45	transferase	transferase

transport	transport protein
tubulin	tubulin
ubiquitin	ubiquitin
unclassified	Protein not categorized into one of the aforementioned protein families
5	water channel protein
water channel	

Column 6 of Table 1, entitled, "Tissue Expression", denotes tissues, represented by four-digit numbers, in which RNA segments giving rise to the SeqCalling™ fragments used to assemble each ORF nucleic acid sequences is present. Tissues or cells corresponding to the numbers are provided in Table 2.

- 5 Table 2 provides generally a correspondence between tissues and diseases or pathologies related to the tissue. Column 1 of Table 2, entitled "tissue id", provides the tissue identification number used in Column 6 of Table 1. The tissue id number runs serially from 1000 to 1058. Column 2 of Table 2, entitled "tissue hierarchy", identifies the tissue and a larger tissue or organ system identified by the identification number of Column 1. Column 3 of Table 2, entitled
10 "Common conditions/diseases", and Column 4 of Table 2, entitled "Other diseases", provide respectively a group of principal diseases, pathologies or conditions, and a group of additional diseases, pathologies or conditions, related to the tissue named in Column 2.

- 15 ORFX nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various ORFX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families indicated in Table 1, and/or according to the presence of domains and sequence relatedness to previously described proteins as summarized in Table 1.

- 20 ORFX nucleic acids and polypeptides according to the invention can also be used to identify cell types listed in Table 1 for an indicated ORFX according to the invention. Additional utilities for ORFX nucleic acids and polypeptides according to the invention are disclosed herein.

ORFX Nucleic Acids

- The novel nucleic acids of the invention include those that encode an ORFX or ORFX-like protein, or biologically active portions thereof, including those in the Sequence Listing, which are denoted SEQ ID NOS. 1-22982. In one embodiment, the nucleic acids include nucleic acids
25 encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. In one embodiment, the encoded polypeptides can thus include, *e.g.*, the amino acid

sequences of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744 includes the nucleic acid sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or a fragment thereof. Additionally, another embodiment of the invention includes nucleic acids that are mutants or variants of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its ORFX -like activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, including fragments, derivatives, analogs and homolog thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ORFX-encoding nucleic acids (*e.g.*, ORFX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of ORFX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

- An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated ORFX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.
- 15 A nucleic acid molecule of any nucleotide sequence within SEQ ID NOS: 1-22982, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. In one embodiment, using all or a portion of the nucleic acid sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 as a hybridization probe, ORFX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)
- 20 A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to

ORFX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term “oligonucleotide” refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction.

- 5 A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than
10 100 nt in length would further comprise at least 6 contiguous nucleotides of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

- In another embodiment, an isolated nucleic acid molecule of the invention comprises a
15 nucleic acid molecule that is a complement of the nucleotide sequence shown in any of the DNA sequences of SEQ ID NO. 1-22982, including any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO. 4987, 4989, 4991, 20 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 that it can hydrogen bond with little or no
25 mismatches to the nucleotide sequence shown in of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, thereby forming a stable duplex.

- As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated
30

polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another 5 polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of the nucleic acids of SEQ ID NO 1-22982, more specifically SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, e.g., a fragment that can be used as a probe or primer, or a fragment 10 encoding a biologically active portion of ORFX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full 15 length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but 20 not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the 25 invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under 30 stringent, moderately stringent, or low stringent conditions. See e.g. Ausubel, *et al.*, CURRENT

PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489,

5 which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of ORFX polypeptide. Isoforms can be expressed in different tissues of the 10 same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a ORFX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, e.g., mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not 15 limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human ORFX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of, e.g., SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 20 19478, 11740, 11742 and 11744 as well as a polypeptide having ORFX activity. Biological activities of the ORFX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human ORFX polypeptide.

The nucleotide sequence determined from the cloning of the human ORFX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed 25 and/or cloning ORFX homologues in other cell types, e.g., from other tissues, as well as ORFX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of, e.g., SEQ ID NO. 4987, 4989, 30 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743;

or an anti-sense strand nucleotide sequence of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743; or of a naturally occurring mutant of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743.

- 5 Probes based on the human ORFX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ORFX protein, such as by
10 measuring a level of a ORFX-encoding nucleic acid in a sample of cells from a subject e.g., detecting ORFX mRNA levels or determining whether a genomic ORFX gene has been mutated or deleted.

“A polypeptide having a biologically active portion of ORFX” refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the
15 present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a “biologically active portion of ORFX” can be prepared by isolating a portion of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, that encodes a polypeptide having a ORFX biological activity (biological activities of the ORFX proteins are
20 summarized in Table 1), expressing the encoded portion of ORFX protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ORFX. For example, a nucleic acid fragment encoding a biologically active portion of ORFX can optionally include a domain as shown in Table 1, column 4.

25 **ORFX variants**

The invention further encompasses nucleic acid molecules that differ from the disclosed ORFX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ORFX protein as that encoded by the nucleotide sequence shown in, any of the nucleic acid sequences of SEQ ID NO: 1-22982, including, for example, SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and
30 11743.

11743. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of the amino acid sequences of SEQ ID NO: 1-22982, including, for example, SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743.

5 In addition to the human ORFX nucleotide sequence shown in any of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of ORFX may exist within a population (e.g., the human population). Such genetic polymorphism in the ORFX gene may 10 exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ORFX protein, preferably a mammalian ORFX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ORFX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in ORFX that are 15 the result of natural allelic variation and that do not alter the functional activity of ORFX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ORFX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 20 11741 and 11743, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the ORFX cDNAs of the invention can be isolated based on their homology to the human ORFX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

25 In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In 30 another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding

region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ORFX proteins derived from species other than 5 human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other 10 sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the 15 probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides 20 (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, 25 the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more 30 washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that

hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2 n -1 (wherein $n = 1$ to 11482) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

- 5 In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X
10 SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.
15 In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2 n -1 (wherein $n = 1$ to 11491), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100
20 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND
25 EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative mutations

- In addition to naturally-occurring allelic variants of the ORFX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by
30 mutation into the nucleotide sequences of any of SEQ ID NO: 1-22982, and more specifically

into the nucleotide sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, thereby leading to changes in the amino acid sequence of the encoded ORFX protein, without altering the functional ability of the ORFX protein. For example, in a preferred embodiment, nucleotide substitutions leading to 5 amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ORFX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid 10 residues that are conserved among the ORFX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Amino acid residues that are conserved among members of an ORFX family members are predicted to be less amenable to alteration. For example, an ORFX protein according to the present invention can contain at least one domain (*e.g.*, as shown in Table 1) that is a typically 15 conserved region in an ORFX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ORFX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding ORFX 20 proteins that contain changes in amino acid residues that are not essential for activity. In one embodiment, such ORFX proteins differ in amino acid sequence from any of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an 25 amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, more preferably at least about 90%, 95%,

98%, and most preferably at least about 99% homologous to SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743.

An isolated nucleic acid molecule encoding a ORFX protein homologous to the protein of any of, e.g., SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence, SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743.

Mutations can be introduced into SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ORFX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ORFX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ORFX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2n-1 (wherein n = 1 to 11491), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant ORFX protein can be assayed for (1) the ability to form protein:protein interactions with other ORFX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ORFX protein and a ORFX

receptor; (3) the ability of a mutant ORFX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind BRA protein; or (5) the ability to specifically bind an anti-ORFX protein antibody.

Antisense

5 Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of any nucleotide sequence of SEQ ID NO: 1-22982, specifically SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a
10 nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ORFX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments,
15 homologs, derivatives and analogs of a ORFX protein of any of, *e.g.*, SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744 or antisense nucleic acids complementary to a ORFX nucleic acid sequence of, *e.g.*, SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 are additionally provided.

20 In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of a human ORFX that corresponds to any of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 25 11742 and 11744. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ORFX disclosed herein, *e.g.*, SEQ ID NO. 30 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739,

11741 and 11743, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ORFX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ORFX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ORFX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or

- genomic DNA encoding a ORFX protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the
- 5 double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic
- 10 acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.
- 15 In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'- α -methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 20 327-330).

Ribozymes and PNA moieties

- Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are
- 25 carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a

30 single-stranded nucleic acid, such as an mRNA, to which they have a complementary region.

Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ORFX mRNA transcripts to thereby inhibit translation of ORFX mRNA. A ribozyme having specificity for a ORFX-encoding nucleic acid can be designed based upon the nucleotide sequence of a ORFX DNA disclosed herein, e.g.,

5 SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ORFX-encoding mRNA. See, e.g., Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, ORFX mRNA can be used to

10 select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, ORFX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the ORFX (e.g., the ORFX promoter and/or enhancers) to form triple helical structures that prevent transcription of the ORFX gene in target cells. See

15 generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of ORFX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be

20 modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic

25 strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ORFX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigenic agents for sequence-specific modulation of gene

30 expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of

ORFX can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

- 5 In another embodiment, PNAs of ORFX can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ORFX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g.,
- 10 RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For
- 15 example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996)
- 20 above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-1124.

- In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556;
- 25 Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, e.g., Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, e.g., Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a

peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

ORFX polypeptides

The novel protein of the invention includes the ORFX-like protein whose sequence is
5 provided in any of the amino acid sequences in SEQ ID NO: 1-22982, including SEQ ID NO:
4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740,
11742 and 11744. The invention also includes a mutant or variant protein any of whose residues
may be changed while still encoding a protein that maintains its ORFX-like activities and
physiological functions, or a functional fragment thereof. For example, the invention includes the
10 polypeptides encoded by the variant ORFX nucleic acids described above. In the mutant or
variant protein, up to 20% or more of the residues may be so changed.

In general, an ORFX -like variant that preserves ORFX-like function includes any variant
in which residues at a particular position in the sequence have been substituted by other amino
acids, and further include the possibility of inserting an additional residue or residues between two
15 residues of the parent protein as well as the possibility of deleting one or more residues from the
parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the
invention. In favorable circumstances, the substitution is a conservative substitution as defined
above. Furthermore, without limiting the scope of the invention, positions of any of, e.g., SEQ ID
NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740,
20 11742 and 11744 may be substitute such that a mutant or variant protein may include one or more
substitutions

The invention also includes isolated ORFX proteins, and biologically active portions
thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide
fragments suitable for use as immunogens to raise anti-ORFX antibodies. In one embodiment,
25 native ORFX proteins can be isolated from cells or tissue sources by an appropriate purification
scheme using standard protein purification techniques. In another embodiment, ORFX proteins
are produced by recombinant DNA techniques. Alternative to recombinant expression, a ORFX
protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially
30 free of cellular material or other contaminating proteins from the cell or tissue source from which

the ORFX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ORFX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language 5 "substantially free of cellular material" includes preparations of ORFX protein having less than about 30% (by dry weight) of non-ORFX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ORFX protein, still more preferably less than about 10% of non-ORFX protein, and most preferably less than about 5% non-ORFX protein. When the ORFX protein or biologically active portion thereof is recombinantly 10 produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes 15 preparations of ORFX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein having less than about 30% (by dry weight) of chemical precursors or non-ORFX chemicals, more preferably less than about 20% chemical precursors or non-ORFX chemicals, still more preferably less than about 10% chemical precursors or non-ORFX chemicals, and most 20 preferably less than about 5% chemical precursors or non-ORFX chemicals.

Biologically active portions of a ORFX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ORFX protein, that include fewer amino acids than the full length ORFX proteins, and exhibit at least one activity of a ORFX protein. Typically, biologically active portions comprise a domain or 25 motif with at least one activity of the ORFX protein. A biologically active portion of a ORFX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ORFX protein of the present invention may contain at least one of the above-identified domains conserved between the FGF family of proteins. Moreover, other biologically active portions, in which other regions of the protein are deleted, can

be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ORFX protein.

In an embodiment, the ORFX protein has an amino acid sequence shown in any of the amino acid sequences of SEQ ID NO: 1-22982, including any of SEQ ID NO: 4988, 4990, 4992, 5 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744. In other embodiments, the ORFX protein is substantially homologous to, e.g., any of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 10 11742 and 11744 and retains the functional activity of the protein of any of, e.g., SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the ORFX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of, e.g., SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 15 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744 and retains the functional activity of the ORFX proteins of the corresponding polypeptide having the sequence of SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744.

Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or 25 nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, 30 such as GAP software provided in the GCG program package. See, *Needleman and Wunsch* 1970

J Mol Biol 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) 5 part of the DNA sequence shown in SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two 10 optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term 15 "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally 20 aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

25 **Chimeric and fusion proteins**

The invention also provides ORFX chimeric or fusion proteins. As used herein, a ORFX "chimeric protein" or "fusion protein" includes a ORFX polypeptide operatively linked to a non-ORFX polypeptide. A "ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ORFX, whereas a "non-ORFX polypeptide" refers to a polypeptide 30 having an amino acid sequence corresponding to a protein that is not substantially homologous to

the ORFX protein, *e.g.*, a protein that is different from the ORFX protein and that is derived from the same or a different organism. Within a ORFX fusion protein the ORFX polypeptide can correspond to all or a portion of a ORFX protein. In one embodiment, a ORFX fusion protein comprises at least one biologically active portion of a ORFX protein. In another embodiment, a

5 ORFX fusion protein comprises at least two biologically active portions of a ORFX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the ORFX polypeptide and the non-ORFX polypeptide are fused in-frame to each other. The non-ORFX polypeptide can be fused to the N-terminus or C-terminus of the ORFX polypeptide.

For example, in one embodiment a ORFX fusion protein comprises a ORFX polypeptide

10 operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate ORFX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-ORFX fusion protein in which the

15 ORFX sequences are fused to the C-terminus of the GST (*i.e.*, glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ORFX.

In yet another embodiment, the fusion protein is a ORFX protein containing a heterologous signal sequence at its N-terminus. For example, the native ORFX signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of ORFX can be increased through use

20 of a heterologous signal sequence.

In another embodiment, the fusion protein is a ORFX-immunoglobulin fusion protein in which the ORFX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The ORFX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject

25 to inhibit an interaction between a ORFX ligand and a ORFX protein on the surface of a cell, to thereby suppress ORFX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated ORFX ligand of the invention is an ORFX receptor. The ORFX-immunoglobulin fusion proteins can be used to modulate the bioavailability of a ORFX cognate ligand. Inhibition of the ORFX ligand/ORFX interaction may be useful therapeutically for both the treatment of

30 proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting)

cell survival. Moreover, the ORFX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ORFX antibodies in a subject, to purify ORFX ligands, and in screening assays to identify molecules that inhibit the interaction of ORFX with a ORFX ligand.

- A ORFX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A ORFX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ORFX protein.

ORFX agonists and antagonists

- The present invention also pertains to variants of the ORFX proteins that function as either ORFX agonists (mimetics) or as ORFX antagonists. Variants of the ORFX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ORFX protein. An agonist of the ORFX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ORFX protein. An antagonist of the ORFX protein can inhibit one or more of the activities of the naturally occurring form of the ORFX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ORFX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ORFX proteins.

Variants of the ORFX protein that function as either ORFX agonists (mimetics) or as ORFX antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the ORFX protein for ORFX protein agonist or antagonist activity. In one embodiment, a variegated library of ORFX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ORFX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ORFX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of ORFX sequences therein. There are a variety of methods which can be used to produce libraries of potential ORFX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ORFX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

Polypeptide libraries

In addition, libraries of fragments of the ORFX protein coding sequence can be used to generate a variegated population of ORFX fragments for screening and subsequent selection of variants of a ORFX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a ORFX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ORFX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene

- products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of ORFX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ORFX variants (Arkin and Yourvan (1992) 5 PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).
- 10 PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

Anti-ORFX Antibodies

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

- An isolated ORFX protein, or a portion or fragment thereof, can be used as an immunogen 15 to generate antibodies that bind ORFX using standard techniques for polyclonal and monoclonal antibody preparation. Full-length ORFX protein can be used. Alternatively, the invention provides antigenic peptide fragments of ORFX for use as immunogens. The antigenic peptide of ORFX comprises at least 4 amino acid residues of the amino acid sequence shown in any of SEQ ID NO:2n (wherein n = 1 to 11491). The antigenic peptide encompasses an epitope of ORFX 20 such that an antibody raised against the peptide forms a specific immune complex with ORFX. The antigenic peptide may comprise at least 6 aa residues, at least 8 aa residues, at least 10 aa residues, at least 15 aa residues, at least 20 aa residues, or at least 30 aa residues. In one embodiment of the invention, the antigenic peptide comprises a polypeptide comprising at least 6 contiguous amino acids of any of the amino acid sequences of SEQ ID NO: 1-22982, including, 25 SEQ ID NO: 4988, 4990, 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744.

- In an embodiment of the invention, epitopes encompassed by the antigenic peptide are regions of ORFX that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and 30 hydrophobicity may be generated by any method well known in the art, including, for example,

the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, e.g., Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

As disclosed herein, an ORFX protein sequence of any of, e.g., SEQ ID NO: 4988, 4990, 5 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744, or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that 10 specifically binds (immunoreacts with) an antigen, such as ORFX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and F_{(ab)2} fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human ORFX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a ORFX protein sequence of any of, e.g., SEQ ID NO: 4988, 4990, 15 4992, 4994, 4996, 4998, 5000, 5076, 5078, 5080, 19474, 19476, 19478, 11740, 11742 and 11744 or derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic 20 preparation can contain, for example, recombinantly expressed ORFX protein or a chemically synthesized ORFX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), 25 human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ORFX can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, 30 refers to a population of antibody molecules that contain only one species of an antigen binding

site capable of immunoreacting with a particular epitope of ORFX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ORFX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ORFX protein, or derivatives, fragments, analogs or homologs thereof, any technique that

5 provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER

10 THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations are incorporated herein by reference in

15 their entirety

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ORFX protein (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, *et al.*, 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a ORFX protein or derivatives, fragments, analogs or homologs thereof.

20 Non-human antibodies can be "humanized" by techniques well known in the art. See e.g., U.S. Patent No. 5,225,539. Each of the above citations are incorporated herein by reference. Antibody fragments that contain the idiotypes to a ORFX protein may be produced by techniques known in the art including, but not limited to: (i) an F_(ab')2 fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_(ab')2 fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

25 Additionally, recombinant anti-ORFX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such

chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 5 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988), *J. Natl Cancer Inst* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. 10 No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060. Each of the above citations are incorporated herein by reference.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other 15 immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ORFX protein is facilitated by generation of hybridomas that bind to the fragment of a ORFX protein possessing such a domain. Antibodies that are specific for one or more domains within a ORFX protein, *e.g.*, the domain spanning the first fifty amino-terminal residues specific to ORFX when compared to FGF-9, or 20 derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ORFX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ORFX protein (*e.g.*, for use in measuring levels of the ORFX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for ORFX proteins, or 25 derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ORFX antibody (*e.g.*, monoclonal antibody) can be used to isolate ORFX by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ORFX antibody can facilitate the purification of natural ORFX from cells and of recombinantly produced 30 ORFX expressed in host cells. Moreover, an anti-ORFX antibody can be used to detect ORFX

protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ORFX protein. Anti-ORFX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

ORFX Recombinant Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ORFX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*,

replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., ORFX proteins, mutant forms of ORFX, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ORFX in prokaryotic or eukaryotic cells. For example, ORFX can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-

fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ORFX expression vector is a yeast expression vector.

Examples of vectors for expression in yeast *S. cerevisiae* include pYEpSec1 (Baldari, *et al.*, (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, ORFX can be expressed in insect cells using baculovirus expression vectors.

Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells)

include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include 5 pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, 10 MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are 15 known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific 20 promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein 25 promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to 30 ORFX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense

orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or 5 attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

10 Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such 15 progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ORFX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in 20 the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride 25 co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the 30 expression vector and transfection technique used, only a small fraction of cells may integrate the

foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a 5 selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be 10 used to produce (*i.e.*, express) ORFX protein. Accordingly, the invention further provides methods for producing ORFX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding ORFX has been introduced) in a suitable medium such that ORFX protein is produced. In another embodiment, the method further comprises isolating ORFX from the 15 medium or the host cell.

Transgenic animals

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which ORFX-coding sequences have been introduced. Such host cells 20 can then be used to create non-human transgenic animals in which exogenous ORFX sequences have been introduced into their genome or homologous recombinant animals in which endogenous ORFX sequences have been altered. Such animals are useful for studying the function and/or activity of ORFX and for identifying and/or evaluating modulators of ORFX activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, 25 more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more 30 cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant

"animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous ORFX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

5 A transgenic animal of the invention can be created by introducing ORFX-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human ORFX DNA sequence of SEQ ID NO:2n-1 (wherein n = 1 to 11482) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of
10 the human ORFX gene, such as a mouse ORFX gene, can be isolated based on hybridization to the human ORFX cDNA (described further above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the ORFX transgene to direct expression of ORFX protein to particular cells. Methods for
15 generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan 1986, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified
20 based upon the presence of the ORFX transgene in its genome and/or expression of ORFX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding ORFX can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a
25 portion of a ORFX gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the ORFX gene. The ORFX gene can be a human gene SEQ ID NO. 4987, 4989, 4991, 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743, but more preferably, is a non-human homologue of a human ORFX gene. For example, a mouse homologue of human ORFX gene of SEQ ID NO. 4987, 4989, 4991,
30 4993, 4995, 4997, 4999, 5075, 5077, 5079, 19473, 19475, 19477, 11739, 11741 and 11743 can be

used to construct a homologous recombination vector suitable for altering an endogenous ORFX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous ORFX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

5 Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous ORFX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous ORFX protein). In the homologous recombination vector, the altered portion of the ORFX gene is flanked at its 5' and 3' ends by additional nucleic acid of the ORFX gene to allow for
10 homologous recombination to occur between the exogenous ORFX gene carried by the vector and an endogenous ORFX gene in an embryonic stem cell. The additional flanking ORFX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector. See *e.g.*, Thomas *et al.* (1987) *Cell* 51:503 for a description of homologous
15 recombination vectors. The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced ORFX gene has homologously recombined with the endogenous ORFX gene are selected (see *e.g.*, Li *et al.* (1992) *Cell* 69:915).

20 The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. See *e.g.*, Bradley 1987, In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous
25 recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Curr Opin Biotechnol* 2:823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

30 In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP

recombinase system, see, e.g., Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

5 Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813. In brief, a cell, 10 e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female 15 foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Pharmaceutical Compositions

The ORFX nucleic acid molecules, ORFX proteins, and anti-ORFX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for 20 administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of 25 Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any 30 conventional media or agent is incompatible with the active compound, use thereof in the

compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

- A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.
- 15 Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by

including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ORFX protein or anti-ORFX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or

suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

5 In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be
10 apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

15 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for
20 the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by any of a number of routes, *e.g.*, as described in U.S. Patent Nos. 5,703,055. Delivery can thus also include, *e.g.*, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the
30 pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Additional Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein
5 can be used in one or more of the following methods: (a) screening assays; (b) detection assays
(e.g., chromosomal mapping, cell and tissue typing, forensic biology), (c) predictive medicine
(e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and
(d) methods of treatment (e.g., therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used to express ORFX protein
10 (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect
ORFX mRNA (e.g., in a biological sample) or a genetic lesion in a ORFX gene, and to modulate
ORFX activity, as described further below. In addition, the ORFX proteins can be used to screen
drugs or compounds that modulate the ORFX activity or expression as well as to treat disorders
characterized by insufficient or excessive production of ORFX protein, for example proliferative
15 or differentiative disorders, or production of ORFX protein forms that have decreased or aberrant
activity compared to ORFX wild type protein. In addition, the anti-ORFX antibodies of the
invention can be used to detect and isolate ORFX proteins and modulate ORFX activity.

This invention further pertains to novel agents identified by the above described screening
assays and uses thereof for treatments as described herein.

20 Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for
identifying modulators, *i.e.*, candidate or test compounds or agents (e.g., peptides,
peptidomimetics, small molecules or other drugs) that bind to ORFX proteins or have a
stimulatory or inhibitory effect on, for example, ORFX expression or ORFX activity.

25 In one embodiment, the invention provides assays for screening candidate or test
compounds which bind to or modulate the activity of a ORFX protein or polypeptide or
biologically active portion thereof. The test compounds of the present invention can be obtained
using any of the numerous approaches in combinatorial library methods known in the art,
including: biological libraries; spatially addressable parallel solid phase or solution phase

libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam
5 (1997) *Anticancer Drug Des* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc Natl Acad Sci U.S.A.* 90:6909; Erb *et al.* (1994) *Proc Natl Acad Sci U.S.A.* 91:11422; Zuckermann *et al.* (1994) *J Med Chem* 37:2678; Cho *et al.* (1993)
10 *Science* 261:1303; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2059; Carell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2061; and Gallop *et al.* (1994) *J Med Chem* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), on chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner USP '409),
15 plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc Natl Acad Sci U.S.A.* 87:6378-6382; Felici (1991) *J Mol Biol* 222:301-310; Ladner above.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a ORFX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the ORFX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the ORFX protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with
20 ¹²⁵I, ³⁵S, ¹⁴C, or ³H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.
25 In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a known
30

compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or a biologically active portion thereof as compared to the known compound.

5 In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion 10 thereof. Determining the ability of the test compound to modulate the activity of ORFX or a biologically active portion thereof can be accomplished, for example, by determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule. As used herein, a "target molecule" is a molecule with which a ORFX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a ORFX interacting protein, a molecule on the 15 surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A ORFX target molecule can be a non-ORFX molecule or a ORFX protein or polypeptide of the present invention. In one embodiment, a ORFX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a 20 compound to a membrane-bound ORFX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with ORFX.

Determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by one of the methods described above for determining direct 25 binding. In one embodiment, determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the 30 induction of a reporter gene (comprising a ORFX-responsive regulatory element operatively

linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

- In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the ORFX protein or biologically active portion thereof. Binding of the test compound to the ORFX protein can be determined either directly or indirectly as described above. In one embodiment, the assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or biologically active portion thereof as compared to the known compound.

- In another embodiment, an assay is a cell-free assay comprising contacting ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX can be accomplished, for example, by determining the ability of the ORFX protein to bind to a ORFX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of ORFX can be accomplished by determining the ability of the ORFX protein further modulate a ORFX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

- In yet another embodiment, the cell-free assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the ORFX protein to preferentially bind to or modulate the activity of a ORFX target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of ORFX. In the case of cell-free assays comprising the membrane-bound form of ORFX, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of ORFX is maintained in solution. Examples of such solubilizing agents include

- 5 non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl)dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

10 In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either ORFX or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to ORFX, or interaction of ORFX with a target molecule in the presence and absence of a candidate compound, can be accomplished in
15 any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-ORFX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized
20 microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or ORFX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components; the matrix immobilized in the case of beads, complex determined either directly or indirectly, for
25 example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ORFX binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either ORFX or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated ORFX or target molecules can be
30 prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*,

biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with ORFX or target molecules, but which do not interfere with binding of the ORFX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or ORFX trapped in the wells by

- 5 antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the ORFX or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the ORFX or target molecule.

In another embodiment, modulators of ORFX expression are identified in a method
10 wherein a cell is contacted with a candidate compound and the expression of ORFX mRNA or protein in the cell is determined. The level of expression of ORFX mRNA or protein in the presence of the candidate compound is compared to the level of expression of ORFX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of ORFX expression based on this comparison. For example, when
15 expression of ORFX mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ORFX mRNA or protein expression. Alternatively, when expression of ORFX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of ORFX
20 mRNA or protein expression. The level of ORFX mRNA or protein expression in the cells can be determined by methods described herein for detecting ORFX mRNA or protein.

In yet another aspect of the invention, the ORFX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J Biol Chem 268:12046-12054; Bartel *et al.* (1993)
25 Biotechniques 14:920-924; Iwabuchi *et al.* (1993) Oncogene 8:1693-1696; and Brent WO94/10300), to identify other proteins that bind to or interact with ORFX ("ORFX-binding proteins" or "ORFX-bp") and modulate ORFX activity. Such ORFX-binding proteins are also likely to be involved in the propagation of signals by the ORFX proteins as, for example, upstream or downstream elements of the ORFX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for ORFX is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other 5 construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a ORFX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., 10 LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with ORFX.

This invention further pertains to novel agents identified by the above-described screening 15 assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, 20 thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

The ORFX sequences of the present invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for 25 identification. The sequences of the present invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Pat. No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the ORFX sequences described herein can be used to prepare two

PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The ORFX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about 5 once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

10

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to 15 differentiate individuals. The noncoding sequences of SEQ ID NO:2n-1 (wherein $n = 1$ to 11491), as described above, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

20 **Predictive Medicine**

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or 25 nucleic acid expression as well as ORFX activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic 30 acid expression or activity. For example, mutations in a ORFX gene can be assayed in a

biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

- Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)
- 10 Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

Use of Partial ORFX Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the ORFX sequences or portions thereof, *e.g.*, fragments derived

from the noncoding regions of one or more of SEQ ID NO: $2n-1$ (where $n = 1$ to 11491), having a length of at least 20 bases, preferably at least 30 bases.

The ORFX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue, etc. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such ORFX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, ORFX primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells 10 in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one 15 aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether 20 an individual is at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

25 Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics").

Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the 30 individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of ORFX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

5 Other conditions in which proliferation of cells plays a role include tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, Kaposi's sarcoma and rheumatoid arthritis.

An ORFX polypeptide may be used to identify an interacting polypeptide a sample or tissue. The method comprises contacting the sample or tissue with ORFX, allowing formation of 10 a complex between the ORFX polypeptide and the interacting polypeptide, and detecting the complex, if present.

The proteins of the invention may be used to stimulate production of antibodies specifically binding the proteins. Such antibodies may be used in immunodiagnostic procedures to detect the occurrence of the protein in a sample. The proteins of the invention may be used to 15 stimulate cell growth and cell proliferation in conditions in which such growth would be favorable. An example would be to counteract toxic side effects of chemotherapeutic agents on, for example, hematopoiesis and platelet formation, linings of the gastrointestinal tract, and hair follicles. They may also be used to stimulate new cell growth in neurological disorders including, for example, Alzheimer's disease. Alternatively, antagonistic treatments may be administered in 20 which an antibody specifically binding the ORFX -like proteins of the invention would abrogate the specific growth-inducing effects of the proteins. Such antibodies may be useful, for example, in the treatment of proliferative disorders including various tumors and benign hyperplasias.

Polynucleotides or oligonucleotides corresponding to any one portion of the ORFX nucleic acids of SEQ ID NO:2n-1 (wherein $n = 1$ to 11482) may be used to detect DNA 25 containing a corresponding ORF gene, or detect the expression of a corresponding ORFX gene, or ORFX-like gene. For example, an ORFX nucleic acid expressed in a particular cell or tissue, as noted in Table 2, can be used to identify the presence of that particular cell type.

An exemplary method for detecting the presence or absence of ORFX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological

sample with a compound or an agent capable of detecting ORFX protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes ORFX protein such that the presence of ORFX is detected in the biological sample. An agent for detecting ORFX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to ORFX mRNA or genomic DNA. The nucleic acid probe can 5 be, for example, a full-length ORFX nucleic acid, such as the nucleic acid of SEQ ID NO:2n-1 (wherein n = 1 to 11491), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to ORFX mRNA or genomic DNA, as described above. Other suitable probes for use in the diagnostic assays of the invention are described herein.

10 An agent for detecting ORFX protein is an antibody capable of binding to ORFX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or 15 antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, 20 cells and fluids present within a subject. That is, the detection method of the invention can be used to detect ORFX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of ORFX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of ORFX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations 25 and immunofluorescence. *In vitro* techniques for detection of ORFX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of ORFX protein include introducing into a subject a labeled anti-ORFX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

5 In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting ORFX protein, mRNA, or genomic DNA, such that the presence of ORFX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of ORFX protein, mRNA or genomic DNA in the control sample with the presence of ORFX protein, mRNA or genomic DNA in the test sample.

10 The invention also encompasses kits for detecting the presence of ORFX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting ORFX protein or mRNA in a biological sample; means for determining the amount of ORFX in the sample; and means for comparing the amount of ORFX in the sample with a standard. The 15 compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect ORFX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant ORFX expression or 20 activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity in, e.g., proliferative or differentiative disorders such as hyperplasias, tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, or rheumatoid arthritis, etc.; and glia-associated disorders 25 such as cerebral lesions, diabetic neuropathies, cerebral edema, senile dementia, Alzheimer's disease, etc. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant ORFX expression or activity in which a test sample is obtained from a subject and ORFX protein or nucleic acid (e.g., mRNA, genomic 30 DNA) is detected, wherein the presence of ORFX protein or nucleic acid is diagnostic for a

subject having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

5 Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant ORFX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a
10 proliferative disorder, differentiative disorder, glia-associated disorders, etc. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant ORFX expression or activity in which a test sample is obtained and ORFX protein or nucleic acid is detected (*e.g.*, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a
15 disorder associated with aberrant ORFX expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a ORFX gene, thereby determining if a subject with the lesioned gene is at risk for, or suffers from, a proliferative disorder, differentiative disorder, glia-associated disorder, etc. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or
20 absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a ORFX-protein, or the mis-expression of the ORFX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides from a ORFX gene; (2) an addition of one or more nucleotides to a ORFX gene; (3) a substitution of one or more nucleotides of a ORFX gene, (4) a chromosomal
25 rearrangement of a ORFX gene; (5) an alteration in the level of a messenger RNA transcript of a ORFX gene, (6) aberrant modification of a ORFX gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a ORFX gene, (8) a non-wild type level of a ORFX-protein, (9) allelic loss of a ORFX gene, and (10) inappropriate post-translational modification of a ORFX-protein. As
30 described herein, there are a large number of assay techniques known in the art which can be used

for detecting lesions in a ORFX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a 5 polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the ORFX-gene (see Abravaya *et al.* (1995) *Nucl Acids Res* 23:675-682). This method can include the steps of 10 collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a ORFX gene under conditions such that hybridization and amplification of the ORFX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control 15 sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990, *Proc Natl Acad Sci USA* 87:1874-1878), transcriptional amplification system (Kwoh, *et 20 al.*, 1989, *Proc Natl Acad Sci USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *BioTechnology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

25 In an alternative embodiment, mutations in a ORFX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. 30 Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,493,531) can

be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in ORFX can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds 5 or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in ORFX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* above. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by 10 making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

15 In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the ORFX gene and detect mutations by comparing the sequence of the sample ORFX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977) *PNAS* 74:560 or Sanger (1977) *PNAS* 74:5463. It is also contemplated that any of a variety of 20 automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publ. No. WO 94/16101; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Other methods for detecting mutations in the ORFX gene include methods in which 25 protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type ORFX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single- 30 stranded regions of the duplex such as which will exist due to basepair mismatches between the

control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions.

- 5 After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al* (1988) *Proc Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol* 217:286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins 10 that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in ORFX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a 15 probe based on a ORFX sequence, e.g., a wild-type ORFX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify 20 mutations in ORFX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl Acad Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res*

separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen *et al.* (1991) *Trends Genet* 7:5.

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers *et al* (1985) *Nature* 313:495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki *et al.* (1986) *Nature* 324:163; Saiki *et al.* (1989) *Proc Natl Acad. Sci USA* 86:6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini *et al* (1992) *Mol Cell Probes* 6:1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification. See, e.g., Barany (1991) *Proc Natl Acad Sci USA* 88:189. In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence, making it possible to

detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

- The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein,
- 5 which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a ORFX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which ORFX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal

10 cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on ORFX activity (*e.g.*, ORFX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurological, cancer-related or gestational disorders) associated with aberrant ORFX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996, *Clin Exp Pharmacol Physiol*, 23:983-985 and Linder, 1997, *Clin Chem*, 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

30 conditions transmitted as a single factor altering the way drugs act on the body (altered drug

action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is haemolysis after 5 ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and 10 cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for 15 CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite 20 morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content 25 of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a ORFX modulator, such as a 30 modulator identified by one of the exemplary screening assays described herein.

Monitoring Clinical Efficacy

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied in basic drug screening and in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase ORFX gene expression, protein levels, or upregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting decreased ORFX gene expression, protein levels, or downregulated ORFX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease ORFX gene expression, protein levels, or downregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting increased ORFX gene expression, protein levels, or upregulated ORFX activity. In such clinical trials, the expression or activity of ORFX and, preferably, other genes that have been implicated in, for example, a proliferative or neurological disorder, can be used as a "read out" or marker of the responsiveness of a particular cell.

For example, genes, including ORFX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates ORFX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of ORFX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of ORFX or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, nucleic acid, peptidomimetic, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a ORFX protein,

- mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the pre-
5 administration sample with the ORFX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of ORFX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease
10 expression or activity of ORFX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant
15 ORFX expression or activity.

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be
20 utilized include, but are not limited to, (i) a ORFX polypeptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to a ORFX peptide; (iii) nucleic acids encoding a ORFX peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to a ORFX peptide) that are utilized to "knockout" endogenous function of a ORFX peptide by homologous
25 recombination (see, *e.g.*, Capecchi, 1989, *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between a ORFX peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with
30

Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, a ORFX peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

- 5 Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of a ORFX peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS)
- 10 polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, etc.).

- In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant ORFX expression or activity, by administering to the subject an agent that modulates ORFX expression or at least one ORFX activity. Subjects at risk for a disease that is caused or contributed to by aberrant ORFX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the ORFX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of ORFX aberrancy, for example,
- 15 a ORFX agonist or ORFX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

- Another aspect of the invention pertains to methods of modulating ORFX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of ORFX protein activity associated with the cell. An agent that modulates ORFX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a ORFX protein, a peptide, a ORFX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more ORFX protein activity. Examples of such stimulatory agents include active ORFX protein and a nucleic acid molecule encoding ORFX that has been introduced into
- 25 the cell. In another embodiment, the agent inhibits one or more ORFX protein activity. Examples

of such inhibitory agents include antisense ORFX nucleic acid molecules and anti-ORFX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder
5 characterized by aberrant expression or activity of a ORFX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) ORFX expression or activity. In another embodiment, the method involves administering a ORFX protein or nucleic acid molecule as therapy to compensate for reduced or
10 aberrant ORFX expression or activity.

Determination of the Biological Effect of a Therapeutic

In various embodiments of the present invention, suitable *in vitro* or *in vivo* assays are utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

15 In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal
20 model system known in the art may be used prior to administration to human subjects.

Malignancies

Some ORFX polypeptides are expressed in cancerous cells (see, e.g., Tables 1 and 2). Accordingly, the corresponding ORF protein is involved in the regulation of cell proliferation. Accordingly, Therapeutics of the present invention may be useful in the therapeutic or
25 prophylactic treatment of diseases or disorders that are associated with cell hyperproliferation and/or loss of control of cell proliferation (e.g., cancers, malignancies and tumors). For a review of such hyperproliferation disorders, see e.g., Fishman, *et al.*, 1985. MEDICINE, 2nd ed., J.B. Lippincott Co., Philadelphia, PA.

Therapeutics of the present invention may be assayed by any method known within the art for efficacy in treating or preventing malignancies and related disorders. Such assays include, but are not limited to, *in vitro* assays utilizing transformed cells or cells derived from the patient's tumor, as well as *in vivo* assays using animal models of cancer or malignancies. Potentially effective Therapeutics are those that, for example, inhibit the proliferation of tumor-derived or transformed cells in culture or cause a regression of tumors in animal models, in comparison to the controls.

In the practice of the present invention, once a malignancy or cancer has been shown to be amenable to treatment by modulating (*i.e.*, inhibiting, antagonizing or agonizing) activity, that cancer or malignancy may subsequently be treated or prevented by the administration of a Therapeutic that serves to modulate protein function.

Premalignant conditions

The Therapeutics of the present invention that are effective in the therapeutic or prophylactic treatment of cancer or malignancies may also be administered for the treatment of pre-malignant conditions and/or to prevent the progression of a pre-malignancy to a neoplastic or malignant state. Such prophylactic or therapeutic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia or, most particularly, dysplasia has occurred. For a review of such abnormal cell growth see *e.g.*, Robbins & Angell, 1976. *BASIC PATHOLOGY*, 2nd ed., W.B. Saunders Co., Philadelphia, PA.

Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in its structure or function. For example, it has been demonstrated that endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of mature or fully differentiated cell substitutes for another type of mature cell. Metaplasia may occur in epithelial or connective tissue cells. Dysplasia is generally considered a precursor of cancer, and is found mainly in the epithelia. Dysplasia is the most disorderly form of non-neoplastic cell growth, and involves a loss in individual cell uniformity and in the architectural orientation of cells. Dysplasia characteristically

occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively, or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed 5 or malignant phenotype displayed either *in vivo* or *in vitro* within a cell sample derived from a patient, is indicative of the desirability of prophylactic/therapeutic administration of a Therapeutic that possesses the ability to modulate activity of An aforementioned protein. Characteristics of a transformed phenotype include, but are not limited to: (i) morphological changes; (ii) looser substratum attachment; (iii) loss of cell-to-cell contact inhibition; (iv) loss of anchorage 10 dependence; (v) protease release; (vi) increased sugar transport; (vii) decreased serum requirement; (viii) expression of fetal antigens, (ix) disappearance of the 250 kDa cell-surface protein, and the like. See *e.g.*, Richards, *et al.*, 1986. MOLECULAR PATHOLOGY, W.B. Saunders Co., Philadelphia, PA.

In a specific embodiment of the present invention, a patient that exhibits one or more of 15 the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: (i) a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome (*bcr/abl*) for chronic myelogenous leukemia and t(14;18) for follicular lymphoma, etc.); (ii) familial polyposis or Gardner's syndrome (possible forerunners of colon cancer); (iii) monoclonal gammopathy of undetermined significance (a possible precursor of 20 multiple myeloma) and (iv) a first degree kinship with persons having a cancer or pre-cancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, medullary thyroid carcinoma with amyloid production and pheochromocytoma, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, 25 intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia and Bloom's syndrome).

In another embodiment, a Therapeutic of the present invention is administered to a human patient to prevent the progression to breast, colon, lung, pancreatic, or uterine cancer, or melanoma or sarcoma.

Hyperproliferative and dysproliferative disorders

- In one embodiment of the present invention, a Therapeutic is administered in the therapeutic or prophylactic treatment of hyperproliferative or benign dysproliferative disorders. The efficacy in treating or preventing hyperproliferative diseases or disorders of a Therapeutic of 5 the present invention may be assayed by any method known within the art. Such assays include *in vitro* cell proliferation assays, *in vitro* or *in vivo* assays using animal models of hyperproliferative diseases or disorders, or the like. Potentially effective Therapeutics may, for example, promote cell proliferation in culture or cause growth or cell proliferation in animal models in comparison to controls.
- 10 Specific embodiments of the present invention are directed to the treatment or prevention of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes); treatment of keloid (hypertrophic scar) formation causing disfiguring of the skin in which the scarring process interferes with normal renewal; psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination); 15 benign tumors; fibrocystic conditions and tissue hypertrophy (*e.g.*, benign prostatic hypertrophy).

Neurodegenerative disorders

- Some ORFX proteins are found in cell types have been implicated in the deregulation of cellular maturation and apoptosis, which are both characteristic of neurodegenerative disease. Accordingly, Therapeutics of the invention, particularly but not limited to those that modulate (or 20 supply) activity of an aforementioned protein, may be effective in treating or preventing neurodegenerative disease. Therapeutics of the present invention that modulate the activity of an aforementioned protein involved in neurodegenerative disorders can be assayed by any method known in the art for efficacy in treating or preventing such neurodegenerative diseases and disorders. Such assays include *in vitro* assays for regulated cell maturation or inhibition of 25 apoptosis or *in vivo* assays using animal models of neurodegenerative diseases or disorders, or any of the assays described below. Potentially effective Therapeutics, for example but not by way of limitation, promote regulated cell maturation and prevent cell apoptosis in culture, or reduce neurodegeneration in animal models in comparison to controls.

Once a neurodegenerative disease or disorder has been shown to be amenable to treatment by modulation activity, that neurodegenerative disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity. Such diseases include all degenerative disorders involved with aging, especially osteoarthritis and neurodegenerative disorders.

5 **Disorders related to organ transplantation**

Some ORFX can be associated with disorders related to organ transplantation, in particular but not limited to organ rejection. Therapeutics of the invention, particularly those that modulate (or supply) activity, may be effective in treating or preventing diseases or disorders related to organ transplantation. Therapeutics of the invention (particularly Therapeutics that modulate the 10 levels or activity of an aforementioned protein) can be assayed by any method known in the art for efficacy in treating or preventing such diseases and disorders related to organ transplantation. Such assays include *in vitro* assays for using cell culture models as described below, or *in vivo* assays using animal models of diseases and disorders related to organ transplantation, see e.g., below. Potentially effective Therapeutics, for example but not by way of limitation, reduce 15 immune rejection responses in animal models in comparison to controls.

Accordingly, once diseases and disorders related to organ transplantation are shown to be amenable to treatment by modulation of activity, such diseases or disorders can be treated or prevented by administration of a Therapeutic that modulates activity.

Cardiovascular Disease

20 GENX has been implicated in cardiovascular disorders, including in atherosclerotic plaque formation. Diseases such as cardiovascular disease, including cerebral thrombosis or hemorrhage, ischemic heart or renal disease, peripheral vascular disease, or thrombosis of other major vessel, and other diseases, including diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage disease, systemic lupus erythematosus, homocysteinemia, and familial 25 protein or lipid processing diseases, and the like, are either directly or indirectly associated with atherosclerosis. Accordingly, Therapeutics of the invention, particularly those that modulate (or supply) activity or formation may be effective in treating or preventing atherosclerosis-associated diseases or disorders. Therapeutics of the invention (particularly Therapeutics that modulate the

levels or activity) can be assayed by any method known in the art, including those described below, for efficacy in treating or preventing such diseases and disorders.

A vast array of animal and cell culture models exist for processes involved in atherosclerosis. A limited and non-exclusive list of animal models includes knockout mice for premature atherosclerosis (Kurabayashi and Yazaki, 1996, Int. Angiol. 15: 187-194), transgenic mouse models of atherosclerosis (Kappel *et al.*, 1994, FASEB J. 8: 583-592), antisense oligonucleotide treatment of animal models (Callow, 1995, Curr. Opin. Cardiol. 10: 569-576), transgenic rabbit models for atherosclerosis (Taylor, 1997, Ann. N.Y. Acad. Sci 811: 146-152), hypercholesterolemic animal models (Rosenfeld, 1996, Diabetes Res. Clin. Pract. 30 Suppl.: 1-11), hyperlipidemic mice (Paigen *et al.*, 1994, Curr. Opin. Lipidol. 5: 258-264), and inhibition of lipoxygenase in animals (Sigal *et al.*, 1994, Ann. N.Y. Acad. Sci. 714: 211-224). In addition, *in vitro* cell models include but are not limited to monocytes exposed to low density lipoprotein (Frostegard *et al.*, 1996, Atherosclerosis 121: 93-103), cloned vascular smooth muscle cells (Suttles *et al.*, 1995, Exp. Cell Res. 218: 331-338), endothelial cell-derived chemoattractant exposed T cells (Katz *et al.*, 1994, J. Leukoc. Biol. 55: 567-573), cultured human aortic endothelial cells (Farber *et al.*, 1992, Am. J. Physiol. 262: H1088-1085), and foam cell cultures (Libby *et al.*, 1996, Curr Opin Lipidol 7: 330-335). Potentially effective Therapeutics, for example but not by way of limitation, reduce foam cell formation in cell culture models, or reduce atherosclerotic plaque formation in hypercholesterolemic mouse models of atherosclerosis in comparison to controls.

Accordingly, once an atherosclerosis-associated disease or disorder has been shown to be amenable to treatment by modulation of activity or formation, that disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity.

Cytokine and Cell Proliferation/Differentiation Activity

A GENX protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell

proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2,

- 5 CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods: Assays for T-cell or thymocyte proliferation include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 3 and Chapter 7); Takai *et al.*, *J Immunol* 137:3494-10 3500, 1986; Bertagnoli *et al.*, *J Immunol* 145:1706-1712, 1990; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Bertagnolli, *et al.*, *J Immunol* 149:3778-3783, 1992; Bowman *et al.*, *J Immunol* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described by Kruisbeek and Shevach, In: CURRENT 15 PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1, pp. 3.12.1-14, John Wiley and Sons, Toronto 1994; and by Schreiber, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan eds. Vol 1 pp. 6.8.1-8, John Wiley and Sons, Toronto 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described by Bottomly *et al.*, In: CURRENT PROTOCOLS IN 20 IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto 1991; DeVries *et al.*, *J Exp Med* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 336:690-692, 1988; Greenberger *et al.*, *Proc Natl Acad Sci U.S.A.* 80:2931-2938, 1983; Nordan, In: CURRENT 25 PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.6.1-5, John Wiley and Sons, Toronto 1991; Smith *et al.*, *Proc Natl Acad Sci U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, *et al.* In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta, *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and 30 cytokine production) include, without limitation, those described In: CURRENT PROTOCOLS IN

IMMUNOLOGY. Coligan *et al.*, eds., Greene Publishing Associates and Wiley-Interscience (Chapter 3 Chapter 6, Chapter 7); Weinberger *et al.*, *Proc Natl Acad Sci USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur J Immun* 11:405-411, 1981; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988.

5 **Immune Stimulating or Suppressing Activity**

A GENX protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), *e.g.*, in regulating (up or down) growth 10 and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (*e.g.*, HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, 15 herpesviruses, mycobacteria, Leishmania species., malaria species. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, 20 rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is 25 desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The 30 functions of activated T cells may be inhibited by suppressing T cell responses or by inducing

specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or energy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after 5 exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon re-exposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7), *e.g.*, preventing high level 10 lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule 15 which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding 20 costimulatory signal. Blocking B lymphocyte antigen function in this manner prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to energize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient 25 immunosuppression or tolerance in a subject, it may also be necessary to block the function of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples 30 of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the

immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc Natl Acad Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

5 Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and auto-antibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may 10 reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of auto-antibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The 15 efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., FUNDAMENTAL 20 IMMUNOLOGY, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B 25 lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic vital diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed 30 APCs either expressing a peptide of the present invention or together with a stimulatory form of a

soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the patient. Another method of enhancing anti-vital immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and 5 reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a 10 nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the 15 peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to 20 induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II α chain protein and an MHC class 25 II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can 30 also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte

antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the
5 following methods: Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981;
10 Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bowman *et al.*, *J Virology* 61:1992-1998; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Brown *et al.*, *J Immunol* 153:3079-3092, 1994.

15 Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J Immunol* 144:3028-3033, 1990; and Mond and Brunswick In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, (eds.) Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto 1994.

20 Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *J Immunol* 149:3778-3783, 1992.

25 Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery *et al.*, *J Immunol* 134:536-544, 1995; Inaba *et al.*, *J Exp Med* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079, 1995; Porgador *et al.*, *J Exp Med* 182:255-260, 1995; Nair *et al.*, *J Virol* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J Exp Med*

169:1255-1264, 1989; Bhardwaj *et al.*, *J Clin Investig* 94:797-807, 1994; and Inaba *et al.*, *J Exp Med* 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) 5 include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J Immunol* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Internat J Oncol* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development 10 include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell Immunol* 155: 111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc Nat Acad Sci USA* 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A GENX protein of the present invention may be useful in regulation of hematopoiesis 15 and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, *e.g.* in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy 20 to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders 25 such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal 30 hemoglobinuria), as well as in repopulating the stem cell compartment post

irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the
5 following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those
10 described in: Johansson *et al.* *Cellular Biology* 15:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in:
Methylcellulose colony forming assays, Freshney, In: CULTURE OF HEMATOPOIETIC CELLS.
15 Freshney, *et al.* (eds.) Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y 1994; Hirayama *et al.*, *Proc Natl Acad Sci USA* 89:5907-5911, 1992; McNiece and Briddeli, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al.*, *Exp Hematol* 22:353-359, 1994; Ploemacher, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Spoonceret *et al.*,
20 In: CULTURE OF HEMATOPOIETIC CELLS. Freshhey, *et al.*, (eds.) Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Sutherland, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A GENX protein of the present invention also may have utility in compositions used for
25 bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a
30 protein of the invention may have prophylactic use in closed as well as open fracture reduction

and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

- A protein of this invention may also be used in the treatment of periodontal disease, and in
- 5 other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.)
- 10 mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and

15 other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of

20 congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex*

25 *vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and

30 for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous

system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, 5 such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

10 Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

15 It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

20 A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

25 The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. 30 WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, EPIDERMAL WOUND HEALING, pp. 71-112 (Maibach and Rovee, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Menz, *J. Invest. Dermatol.* 71:382-84 (1978).

5 **Activin/Inhibin Activity**

A GENX protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in 10 heterodimers with a member of the inhibin a family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-b group, may be useful as a fertility inducing 15 therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the 20 following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc Natl Acad Sci USA* 83:3091-3095, 1986.

25 **Chemotactic/Chemokinetic Activity**

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action.

Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

5 A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

10 The activity of a protein of the invention may, among other means, be measured by following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population 15 to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Coligan *et al.*, eds. (Chapter 6.12, MEASUREMENT OF ALPHA AND BETA CHEMOKINES 6.12.1-6.12.28); Taub *et al.* *J Clin Invest* 95:1370-1376, 1995; Lind *et al.* *APMIS* 103:140-146, 1995; Muller *et al.*, *Eur J Immunol* 25: 1744-1748; Gruber *et al.* *J Immunol* 152:5860-5867, 1994; Johnston *et al.*, *J Immunol* 153: 1762-20 1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other 25 hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

The activity of a protein of the invention may, among other means, be measured by the 30 following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet *et al.*, *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al.*, *Thrombosis Res.* 45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

5 **Receptor/Ligand Activity**

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and 10 their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, 15 fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in:
20 CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan, *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai *et al.*, *Proc Natl Acad Sci USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 1989; Stoltenborg *et al.*, *J Immunol Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995.

25 **Anti-Inflammatory Activity**

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory

process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic

5 shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

10 **Tumor Inhibition Activity**

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

20 A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido,

- stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and
- 5 treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.
- 10 Neural disorders in general include Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumors of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury, and other neuropathies, epilepsy, and/or tremors.

15 EQUIVALENTS

From the foregoing detailed description of the specific embodiments of the invention, it should be apparent that particular novel compositions and methods involving nucleic acids, polypeptides, antibodies, detection and treatment have been described. Although these particular embodiments have been disclosed herein in detail, this has been done by way of example for

20 purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made as a matter of routine for a person of ordinary skill in the art to the invention without departing from the spirit and scope of the invention as defined by the claims. Indeed, various modifications of the invention in addition to

25 those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

TABLE 1.

Pages 1-930

Entry	Sequence Id	Protein Similarity	Protein Domain	Protein Classification	Tissue Expression
1	80034180 (5897, 5898)	Novel Protein sim. GBank gil120927 sp P03024 GALR_ECOLI - GALACTOSE OPERON REPRESSOR	Contains protein domain (PF00356) - Bacterial regulatory proteins, lacI family	amylase ^e	1006, 1031
2	78769186 (21593, 21594)	Novel Protein sim. GBank gil3643921 (AF086819) - transcription regulator CelR [Thermomonospora fusca]	Contains protein domain (PF00356) - Bacterial regulatory proteins, lacI family	amylase	1016, 1017, 1026
3	25252417 (5061, 5062)	Novel Protein sim. GBank gil11722870 sp P44329 RBSR_HAEIN - RIBOSE OPERON REPRESSOR		amylase	1003
4	87452791 (15017, 15018)	Novel Protein sim. GBank gil170332 sp P55005 AMYB_MAIZE - BETA-AMYLASE (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE)		amylase	1025
5	79758169 (3785, 3786)	Novel Protein sim. GBank gil1838836 dbj BAA11010 - (D64130) alpha-amylase [Sulfolobus solfataricus]		amylase	1008
6	80502141 (4679, 4680)	Novel Protein sim. GBank gil336775 emb CAA20080 - (AL031155) putative transcriptional regulator [Streptomyces coelicolor]		amylase	1004, 1010, 1012, 1023, 1034
7	78271643 (10933, 10934)	Novel Protein sim. GBank gil728851 sp Q05884 AMY_STRL1 - ALPHA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN GLUCANOHYDROLASE)		amylase	1039
8	78514365 (8317, 8318)	Novel Protein sim. GBank gil36890 p1 [S21354 - hypothetical protein • Streptomyces lividans		amylase ^e	1008, 1023, 1026
9	80592427 (22909, 22910)	Novel Protein sim. GBank gil1345834 sp P48199 CRP_RAT - C-REACTIVE PROTEIN PRECURSOR	Contains protein domain (PF00354) - Pentaxin family	amyloid	1003, 1006, 1016, 1017, 1022, 1036, 1039, 1044, 1058

10	39443817 (4669, 4670)	Novel Protein sim. GBank gi 4539685 gb AAD22173.1 AF09901 - (AF099019) hepatocyte growth factor activator inhibitor type 2 splice variant 1 [Mus musculus]		amyloid	1004, 1044
11	80242642 (5583, 5584)	Novel Protein sim. GBank gi 5295890 emb CAB46014.1 - (AJ242649) plasminogen protein (Rattus norvegicus)	Contains protein domain (PF00051) - Kringle domain	apolipoprotein	1001, 1004, 1006, 1034, 1039
12	79183303 (18647, 18648)	Novel Protein sim. GBank gi 3519541 sp P26644 APOH_RAT - BETA-2- GLYCOPROTEIN I PRECURSOR (APOLIPOPROTEIN H) (APO-H)	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	apolipoprotein	1000, 1003, 1008, 1009, 1022, 1026, 1044
13	21420629 (5349, 5350)	Novel Protein sim. GBank gi 5295890 emb CAB46014.1 - (AJ242649) plasminogen protein (Rattus norvegicus)	Contains protein domain (PF00089) - Trypsin	apolipoprotein	1004, 1006, 1010, 1034, 1058
14	17710870 (1679, 1680)	Novel Protein sim. GBank gi 109577 pir JH0101 - apolipoprotein B-100 - mouse (fragment)		apolipoprotein	1044
15	30491676 (5509, 5510)	Novel Protein sim. GBank gi 109577 pir JH0101 - apolipoprotein B-100 - mouse (fragment)		apolipoprotein	1026
16	8351088 (15203, 15204)	Novel Protein sim. GBank gi 14018 sp P19939 APC1_RAT - APOLIPOPROTEIN C-I PRECURSOR (APO-C-I)		apolipoprotein	1006
17	78727869 (8783, 8784)	Novel Protein sim. GBank gi 14029 sp P06759 APC3_RAT - APOLIPOPROTEIN C-III PRECURSOR (APO- C-III)		apolipoprotein	1008, 1009, 1022, 1026, 1039
18	3026312 (3817, 3818)	Novel Protein sim. GBank gi 1170804 sp P44626 LNT_HAEIN - APOLIPOPROTEIN N-ACYLTRANSFERASE (ALP N-ACYL TRANSFERASE) (COPPER HOMEOSTASIS PROTEIN CUTE)		apolipoprotein	1016
19	78676338 (6815, 6816)	Novel Protein sim. GBank gi 4502165 ref NP_001629.1 pAPOF - apolipoprotein F		apolipoprotein	1026

20	80230526 (815, 816)	Novel Protein sim. GBank g 231574 sp P30362 ASPG_LUPAR - L-ASPARAGINASE (L-ASPARAGINE AMIDOHYDROLASE)		asparaginase	1001, 1004, 1008, 1010, 1024
21	79260356 (1583, 1584)	Novel Protein sim. GBank g 295644 (J04421) - ATPase [Saccharomyces cerevisiae]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase_associated	1003
22	20630761 (19087, 19088)	Novel Protein sim. GBank g 295644 (J04421) - ATPase [Saccharomyces cerevisiae]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase_associated	1022
23	11399341 (21325, 21326)	Novel Protein sim. GBank g 3777495.(J92083) - calcium transporting ATPase [Pichia angusta]	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase_associated	1006
24	9848318 (9841, 9842)	Novel Protein sim. GBank g 388056 emb CAA94236 - (Z70271) Similarity to Yeast E1-E2 ATPase (SW; YEAST); cDNA EST EMBL:D37634 comes from this gene; cDNA EST EMBL:D34653 comes from this gene; cDNA EST EMBL:D34527 comes from this gene; cDNA EST EMBL:D32311 comes from this ge...	Contains protein domain (PF00122) - E1-E2 ATPase	ATPase_associated	1039
25	20617672 (18133, 18134)	Novel Protein sim. GBank g 2668599 (AF012537) - ATPase [Pseudomonas aeruginosa]	Contains protein domain (PF00265) - Thymidine kinases	ATPase_associated	1010
26	16841039 (13961, 13962)	Novel Protein sim. GBank g 1870 sp P13454 DNAA_PSEPU - CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA	Contains protein domain (PF00308) - Bacterial dnaA protein	ATPase_associated	1023
27	80223711 (11417, 11418)	Novel Protein sim. GBank g 586351 sp P38328 AR41_YEAST - PROBABLE ARP2/3 COMPLEX 41 KD SUBUNIT (P41-ARC)	Contains protein domain (PF00400) - WD domain, G-beta repeat	ATPase_associated	1024, 1050
28	25250758 (8495, 8496)	Novel Protein sim. GBank g 31220941 sp Q59207 FIX1_BRAJA - NITROGEN FIXATION PROTEIN FIX1 (E1-E2 TYPE CATION ATPASE FIX1)	Contains protein domain (PF00403) - Heavy-metal-associated domain	ATPase_associated	1044

29	42724092 (13883, 13884)	Novel Protein sim. GBank gi 2921168 (AF025656) - HrCA [Streptomyces albus G]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deoR family	ATPase_associated	1001
30	80251181 (5889, 5890)	Novel Protein sim. GBank gi 586900 sp P37571 MECB_BACSU - NEGATIVE REGULATOR OF GENETIC COMPETENCE MECB	Contains protein domain (PF00495) - Chaperonin clip/A/B	ATPase_associated	1006, 1034
31	83199901 (14999, 15000)	Novel Protein sim. GBank gi 205982 (K03222) - prealbumin [Rattus norvegicus]	Contains protein domain (PF00576) - Transthyretin precursor (formerly prealbumin)	ATPase_associated	1003, 1008, 1009, 1022
32	94239717 (2783, 2784)	Novel Protein sim. GBank gi 1094109 prf 1205237A - CCAAT-binding factor [Saccharomyces cerevisiae]	Contains protein domain (PF00808) - Histone-like transcription factors (CBF/NF-Y) and archaeal histones.	ATPase_associated	1022, 1039, 1044, 1058
33	80486103 (18839, 18840)	Novel Protein sim. GBank gi 418575 sp P32795 YME1_YEAST - YME1 PROTEIN (TAT-BINDING HOMOLOG 11) (OSD1 PROTEIN)	Contains protein domain (PF01434) - Peptidase family M41	ATPase_associated	1003, 1004, 1006, 1008, 1010, 1012, 1022, 1023, 1024, 1026, 1027, 1034, 1039, 1054
34	30658161 (12719, 12720)	Novel Protein sim. GBank gi 3122234 sp O24473 F2B_WHEAT-EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA) (P38)	Contains protein domain (PF01873) - Domain found in IF2B/IF5	ATPase_associated	1026
35	57001671 (19347, 19348)	Novel Protein sim. GBank gi 1077333 pir S51473 - probable membrane protein YLR386w - yeast (Saccharomyces cerevisiae)		ATPase_associated	1008

36	27213004 (19343, 19344)	Novel Protein sim. GBank g 114980 sp P28863 SNAB_MOUSE - BETA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-BETA) (SNAP-ALPHA HOMOLOG) (BRAIN PROTEIN) (147)		ATPase_associated	1026
37	79462534 (18195, 18196)	Novel Protein sim. GBank g 1169973 sp P44403 CLPB_HAEIN - CLPB PROTEIN		ATPase_associated	1034
38	37801202 (8521, 8522)	Novel Protein sim. GBank g 1170852 sp P45701 MA12_RABIT - MANNOSYL-OGLYCOSACCHARIDE ALPHA-1,2-MANNOSIDASE (MAN(9)-ALPHA-MANNOSIDASE)		ATPase_associated	1012
39	20294798 (13503, 13504)	Novel Protein sim. GBank g 1175147 sp P4452 ZNUA_HAEIN - HIGH-AFFINITY ZINC UPTAKE SYSTEM PROTEIN ZNUA		ATPase_associated	1034
40	10983631 (5059, 5060)	Novel Protein sim. GBank g 1176081 sp P44815 YGBB_HAEIN - HYPOTHETICAL PROTEIN H0671		ATPase_associated	1022
41	80477264 (20281, 20282)	Novel Protein sim. GBank g 1176203 sp P46442 YHCM_ECOLI - HYPOTHETICAL 43.1 KD PROTEIN IN RPLM-HHOA INTERGENIC REGION (F375)		ATPase_associated	1012
42	80502613 (7397, 7398)	Novel Protein sim. GBank g 134972 sp P16865 STF2_YEAST - ATPASE STABILIZING FACTOR 15 KD PROTEIN		ATPase_associated	1001, 1004, 1006, 1008, 1009, 1012, 1016, 1017, 1022, 1024, 1034, 1038, 1044, 1054
43	11817500 (9131, 9132)	Novel Protein sim. GBank g 135727 sp P25755 THDF_PSEPU - POSSIBLE THIOPHENE AND FURAN OXIDATION PROTEIN THDF		ATPase_associated	1006
44	80260522 (9527, 9528)	Novel Protein sim. GBank g 147364 (M29384) - unidentified reading frame 1 [Escherichia coli]		ATPase_associated	1010, 1034, 1054

45	27839796 (19305, 19306)	Novel Protein sim. GBank gil1705677 sp P54609 CC48_ARATH - CELL DIVISION CYCLE PROTEIN 48 HOMOLOG		ATPase_associated	1006
46	20466021 (1881, 1882)	Novel Protein sim. GBank gil1705678 sp P54774 CC48_SOYBN - CELL DIVISION CYCLE PROTEIN 48 HOMOLOG (VALOSIN CONTAINING PROTEIN HOMOLOG) (VCP)		ATPase_associated	1010
47	86477064 (2875, 2876)	Novel Protein sim. GBank gil1705922 sp P53532 CLPB_CORGL - CLPB PROTEIN		ATPase_associated	1029
48	11090770 (9471, 9472)	Novel Protein sim. GBank gil1705924 sp P53533 CLPB_SYNTP7 - CLPB PROTEIN		ATPase_associated	1006
49	32444968 (16143, 16144)	Novel Protein sim. GBank gil1709142 sp P32559 MSS1_YEAST - MITOCHONDRIAL GTPASE MSS1 PRECURSOR		ATPase_associated	1029
50	11286412 (4351, 4352)	Novel Protein sim. GBank gil1709801 sp Q01939 PRS8_YEAST - 26S PROTEASE REGULATORY SUBUNIT 8 HOMOLOG (SUG1 PROTEIN) (CIM3 PROTEIN) (TAT-BINDING PROTEIN TBY1)		ATPase_associated	1006
51	79258888 (635, 636)	Novel Protein sim. GBank gil172563 (M20662) - SEC18 gene product [Saccharomyces cerevisiae]		ATPase_associated	1003, 1040
52	46802362 (22327, 22328)	Novel Protein sim. GBank gil172563 (M20662) - SEC18 gene product [Saccharomyces cerevisiae]		ATPase_associated	1034
53	14372317 (16805, 16806)	Novel Protein sim. GBank gil2131333 sp S67642 - hypothetical protein YDL100c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1016

54	82125929 (10515) 10516)	Novel Protein sim. GBank gi 2 131784 pir S64954 - hypothetical protein YLR117c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1001, 1003, 1004, 1006, 1008, 1010, 1016, 1022, 1024, 1026, 1031, 1034, 1039, 1044, 1054
55	80218482 (12417) 12418)	Novel Protein sim. GBank gi 2 132242 pir S61029 - hypothetical protein YPL235w - yeast (Saccharomyces cerevisiae)		ATPase_associated	1004, 1022
56	78764331 (3677, 3678)	Novel Protein sim. GBank gi 2 132688 pir S64942 - probable membrane protein YLR106c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1003, 1017, 1044
57	80429767 (10777) 10778)	Novel Protein sim. GBank gi 2 132688 pir S64942 - probable membrane protein YLR106c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1003, 1008, 1017, 1022, 1024, 1031, 1038, 1039
58	80249041 (12477) 12478)	Novel Protein sim. GBank gi 2 132688 pir S64942 - probable membrane protein YLR106c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1006, 1008, 1026
59	80216443 (13523, 13524)	Novel Protein sim. GBank gi 2 132688 pir S64942 - probable membrane protein YLR106c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1003, 1004, 1006, 1008, 1017, 1022, 1029, 1034
60	79470572 (20901, 20902)	Novel Protein sim. GBank gi 2 132688 pir S64942 - probable membrane protein YLR106c - yeast (Saccharomyces cerevisiae)		ATPase_associated	1022
61	27972092 (16253, 16254)	Novel Protein sim. GBank gi 2 245570 (AF005035) - alpha 1,2-mannosidase [Spodoptera frugiperda]		ATPase_associated	1006
62	7524928 (21567, 21568)	Novel Protein sim. GBank gi 2 29516 (AF025468) - No definiton line found [Caenorhabditis elegans]		ATPase_associated	1058

63	21434033 (5547, 5548)	Novel Protein sim. GBank gj 2492830 sp Q03557 AMDZ YEAST - PROBABLE AMIDASE YMR293C		ATPase_associated	1022
64	29198735 (18927, 18928)	Novel Protein sim. GBank gj 2497139 sp Q04779 YMW5 YEAST - HYPOTHETICAL 78.8 KD PROTEIN IN ABF2- CHL12 INTERGENIC REGION		ATPase_associated	1008
65	79852839 (16035, 16036)	Novel Protein sim. GBank gj 2507541 sp P33643 RLUD_ECOLI - RIBOSOMAL LARGE SUBUNIT PSEUDOOURIDYLATE SYNTHASE D (PSEUDOOURIDYLATE SYNTHASE) (URACIL HYDROLYASE)		ATPase_associated	1044, 1058
66	37035355 (14269, 14270)	Novel Protein sim. GBank gj 2627328 (AF004408) - ThIG [Rhizobium etli]		ATPase_associated	1012
67	25267907 (9879, 9880)	Novel Protein sim. GBank gj 2839918 (AC002291) - similar to "tub" protein gp U82468 2072162 [Arabidopsis thaliana]		ATPase_associated	1044
68	79260355 (1477, 1478)	Novel Protein sim. GBank gj 288538 emb CAA50290 - (X70951) orf1 [Saccharomyces cerevisiae]		ATPase_associated	1003, 1008
69	17683899 (929, 930)	Novel Protein sim. GBank gj 2944187 (AF011336) - putative E1-E2 ATPase [Mus musculus]		ATPase_associated	1017
70	37063321 (19043, 19044)	Novel Protein sim. GBank gj 2944187 (AF011336) - putative E1-E2 ATPase [Mus musculus]		ATPase_associated	1050
71	30265416 (6153, 6154)	Novel Protein sim. GBank gj 295644 (J04421) - ATPase [Saccharomyces cerevisiae]		ATPase_associated	1026
72	30655704 (11319, 11320)	Novel Protein sim. GBank gj 295644 (J04421) - ATPase [Saccharomyces cerevisiae]		ATPase_associated	1022
73	27260702 (14105, 14106)	Novel Protein sim. GBank gj 295644 (J04421) - ATPase [Saccharomyces cerevisiae]		ATPase_associated	1022
74	94996320 (15321, 15322)	Novel Protein sim. GBank gj 3328177 (AF074017) - nonsense-mediated mRNA decay trans-acting factor [Caenorhabditis elegans]		ATPase_associated	1022

75	47656560 (17461, 17462), Novel Protein sim. GBank gi 3676746 gb AAC64282.1 AAC6428 - (AF0866332) ARC [Streptomyces coelicolor]		ATPase_associated	1029
76	78296589 (807, 808), Novel Protein sim. GBank gi 3882199 dbj BAA34459.1 - (AB018282) KIAA0739 protein [Homo sapiens]		ATPase_associated	1026, 1036
77	25318650 (3357, 3358), Novel Protein sim. GBank gi 403717 (U49397) - unknown [Streptococcus pyogenes]		ATPase_associated	1008
78	28345053 (6851, 6852), Novel Protein sim. GBank gi 418575 sp P32795 YME1 YEAST - YME1 PROTEIN (TAT-BINDING HOMOLOG 11) (OSD1 PROTEIN)		ATPase_associated	1008, 1012
79	79624339 (13675, 13676), Novel Protein sim. GBank gi 432288 emb CAA79923 - (Z21877) heavy chain of cytoplasmic dynein [Saccharomyces cerevisiae]		ATPase_associated	1001
80	79863926 (20415, 20416), Novel Protein sim. GBank gi 4495126 emb CAB37625.2 - (AL0355583) putative helicase [Schizosaccharomyces pombe]		ATPase_associated	1017
81	21434002 (447, 448), Novel Protein sim. GBank gi 460160 (U06465) - Srv1p [Saccharomyces cerevisiae]		ATPase_associated	1034, 1058
82	39565323 (3347, 3348), Novel Protein sim. GBank gi 465495 sp P33595 YABN_ECOLI - HYPOTHETICAL 63.9 KD PROTEIN IN TBPA- LEUD INTERGENIC REGION (ORF103)		ATPase_associated	1024
83	52390668 (13155, 13156), Novel Protein sim. GBank gi 4678973 emb CAB41297.1 - (AL034345) dJ207H1.1 (axonemal Dynein Heavy Chain protein DNAH) [Homo sapiens]		ATPase_associated	1036
84	35091664 (217, 218), Novel Protein sim. GBank gi 4808343 emb CAB42757.1 - (AL049841) cell division protein ftsH homolog [Streptomyces coelicolor]		ATPase_associated	1050

85	20466909 (13595, 13596)	Novel Protein sim. GBank gi 4808343 emb CAB42757.1 - (AL049841) cell division protein ftsH homolog [Streptomyces caelicolor]		ATPase_associated	1010
86	78785207 (14437, 14438)	Novel Protein sim. GBank gi 5031599 ref NP_005722.1 pARC3 - ARP2/3 protein complex subunit 34		ATPase_associated	1022
87	8489045 (11983, 11984)	Novel Protein sim. GBank gi 5031697 ref NP_005594.1 pFIC1 - familial intrahepatic cholestasis 1, (progressive, Byler disease and benign recurrent)		ATPase_associated	1022, 1024
88	27963785 (10657, 10658)	Novel Protein sim. GBank gi 541748 emb CAA54091 - (X76643) Afg3p [Saccharomyces cerevisiae]		ATPase_associated	1022, 1024, 1034
89	5737021 (5413, 5414)	Novel Protein sim. GBank gi 5457357 emb CAB48398.1 - (Y18314) paraplegin-like protein [Homo sapiens]		ATPase_associated	1058
90	30258244 (12511, 12512)	Novel Protein sim. GBank gi 5457357 emb CAB48398.1 - (Y18314) paraplegin-like protein [Homo sapiens]		ATPase_associated	1026
91	80240719 (16423, 16424)	Novel Protein sim. GBank gi 549723 sp P36062 YKO6_YEAST - HYPOTHETICAL 75.5 KD PROTEIN IN SDH1- CIM5/YTA3 INTERGENIC REGION		ATPase_associated	1001, 1004, 1006, 1010, 1027, 1038
92	11697042 (16777, 16778)	Novel Protein sim. GBank gi 5669564 gb AAD46389.1 AF07047 - (AF070475) NBC-like protein 2 [Rattus norvegicus]		ATPase_associated	1034
93	47652328 (18295, 18296)	Novel Protein sim. GBank gi 585159 sp P37476 FTSH_BACSU - CELL DIVISION PROTEIN FTSH HOMOLOG		ATPase_associated	1029
94	21644312 (11733, 11734)	Novel Protein sim. GBank gi 687208 U03976 - dynein heavy chain iso type 5C [Tripneustes gratilla]		ATPase_associated	1034
95	11815336 (17333, 17334)	Novel Protein sim. GBank gi 3767582 dbj BAA33850 - (ABD18531) dtsR2 [Corynebacterium glutamicum]	biotindep	1006	

96	9892113 (4755, 4756)	Novel Protein sim. GBank gi 4981240 gb AAD35798.1 AE00174 - (AE001742) propionyl-CoA carboxylase, beta subunit [Thermotoga maritima]		biotindep	1034
97	25257259 (2163, 2164)	Novel Protein sim. GBank gi 538966 pir A48665 - methylmalonyl-CoA carboxyltransferase (EC 2.1.3.1) 12S chain - Propionibacterium freudenreichii subsp. shermanii		biotindep	1003, 1026
98	71314225 (19995, 19996)	Novel Protein sim. GBank gi 1352494 spi P49134 ITB1_RAT - FIBRONECTIN RECEPTOR BETA SUBUNIT PRECURSOR (INTEGRIN BETA-1)	Contains protein domain (PF00008) - EGF-like domain	cadherin	1000
99	85547422 (5699, 5700)	Novel Protein sim. GBank gi 4504383 ref NP_001519.1 pHGFA - HGF activator	Contains protein domain (PF00008) - EGF-like domain	cadherin	1053
100	49504200 (19197, 19198)	Novel Protein sim. GBank gi 345960 spi P33450 FAT_DROME - CADHERIN RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN)		cadherin	1040
101	80256978 (4123, 4124)	Novel Protein sim. GBank gi 3122226 1 spi Q12522 IF6_YEAST - EUKARYOTIC TRANSLATION INITIATION FACTOR 6 (EIF-6)		cadherin	1008, 1012, 1022
102	95001848 (4335, 4336)	Novel Protein sim. GBank gi 312226 1 spi Q12522 IF6_YEAST - EUKARYOTIC TRANSLATION INITIATION FACTOR 6 (EIF-6)		cadherin	1006, 1008, 1022, 1038
103	6847376 (19161, 19162)	Novel Protein sim. GBank gi 416742 spi P33146 CADF_MOUSE - MUSCLE-CADHERIN PRECURSOR (M-CADHERIN) (CADHERIN-15) (CADHERIN-14)		cadherin	1022
104	78674409 (9157, 9158)	Novel Protein sim. GBank gi 450563 1 ref NP_002578.1 pPCDH - protocadherin 1 (cadherin-like 1)		cadherin	1009, 1036, 1038

105	65473912 (1273, 1274)	Novel Protein sim. GBank gj 5456982 gb AAD43732.1 - (AF152338) protocadherin gamma C4 [Homo sapiens]		cadherin	1054
106	54841747 (19571, 19572)	Novel Protein sim. GBank gj 1710177 (U76716) - voltage-sensitive calcium channel alpha 1 A [Mus musculus]		calcium_channel	1009
107	20711847 (13117, 13118)	Novel Protein sim. GBank gj 118333 sp P23234 DCIP_ENTCL - INDOLE-3-PYRUVATE DECARBOXYLASE (INDOLEPYRUVATE DECARBOXYLASE)	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1022
108	11093114 (10781, 10782)	Novel Protein sim. GBank gj 118389 sp P26263 DCP3_YEAST - PYRUVATE DECARBOXYLASE ISOZYME 3	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1022
109	80058824 (13631, 13632)	Novel Protein sim. GBank gj 2144527 pir DCBYP - Pyruvate decarboxylase (EC 4.1.1.1) 1 - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1009, 1017, 1022, 1039
110	18414190 (7045, 7046)	Novel Protein sim. GBank gj 2501334 sp Q07471 THI3_YEAST - THIAMINE METABOLISM REGULATORY PROTEIN THI3	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1024
111	29359119 (19983, 19984)	Novel Protein sim. GBank gj 2301334 sp Q07471 THI3_YEAST - THIAMINE METABOLISM REGULATORY PROTEIN THI3	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1006, 1022, 1023
112	85522997 (19787, 19788)	Novel Protein sim. GBank gj 2916912 emb CAA17659.1 - (AL0222004) pdc [Mycobacterium tuberculosis]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1010, 1034, 1049
113	11803584 (5717, 5718)	Novel Protein sim. GBank gj 515236 pdbj 1PYDA - Saccharomyces bayanus	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	carboxylase	1004
114	10887144 (21111, 21112)	Novel Protein sim. GBank gj 1168278 sp P46401 BCCA_MYCTU - ACETYL-PROPYONYL-COENZYME A CARBOXYLASE ALPHA CHAIN [CONTAINS: BIOTIN CARBOXYLASE; BIOTIN CARBOXYL CARRIER PROTEIN (BCCP)]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSSase)	carboxylase	1027

115	3177046 (6371, 6372)	Novel Protein sim. GBank gil1168279 spl P43873 ACCC_HAEIN - BIOTIN CARBOXYLASE (A SUBUNIT OF ACETYL-COA CARBOXYLASE (AC))	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1029
116	13076326 (18855, 18856)	Novel Protein sim. GBank gil1168279 spl P43873 ACCC_HAEIN - BIOTIN CARBOXYLASE (A SUBUNIT OF ACETYL-COA CARBOXYLASE (AC))	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1031
117	80219329 (14409, 14410)	Novel Protein sim. GBank gil171504 (M92156) - acetyl-CoA carboxylase [Saccharomyces cerevisiae]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1022
118	11870440 (16327, 16328)	Novel Protein sim. GBank gil2623051 (AE000942) - biotin carboxylase [Methanobacterium thermoautotrophicum]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1022
119	1987204 (16195, 16196)	Novel Protein sim. GBank gil2983895 (AE000744) - pyruvate carboxylase n-terminal domain [Aquifex aeolicus]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1034
120	11073114 (11721, 11722)	Novel Protein sim. GBank gil471922 gb AAD28553.1 AF11360 - (AF113603) putative acyl-CoA carboxylase complex A subunit [Streptomyces coelicolor]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	carboxylase	1034
121	17906395 (3895, 3896)	Novel Protein sim. GBank gil115577 spl P29193 CAP1_SACHY - PHOSPHOENOLPYRUVATE CARBOXYLASE, HOUSEKEEPING ISOZYME (PEPCASE)	Contains protein domain (PF00311) - Phosphoenolpyruvate carboxylase	carboxylase	1003
122	29210483 (22487, 22488)	Novel Protein sim. GBank gil115587 spl P29195 CAP1_SORVU - PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (PEPCASE) (CP21)	Contains protein domain (PF00311) - Phosphoenolpyruvate carboxylase	carboxylase	1044

123	14984487 (18185, 18186)	Novel Protein sim. GBank gil 1345625 sp P48445 BPL1_YEAST - BIOTIN--PROTEIN LIGASE (BIOTIN--[METHYLMALONYL-COA-CARBOXYL TRANSFERASE] LIGASE / BIOTIN--[PROPIONYL-COA-CARBOXYLASE (ATP-HYDROLYSING)] LIGASE (HOLOCARBOXYLASE SYNTHETASE) (HCS) / BIOTIN--METH...)		carboxylase	1024
124	25132921 (11923, 11924)	Novel Protein sim. GBank gil 171504 (M92156) - acetyl-CoA carboxylase [Saccharomyces cerevisiae]		carboxylase	1003
125	16307319 (12587, 12588)	Novel Protein sim. GBank gil 172102 (J03889) - pyruvate carboxylase (EC 6.4.1.1) [Saccharomyces cerevisiae]		carboxylase	1024
126	79246475 (4899, 4900)	Novel Protein sim. GBank gil 1755924 emb CAA68941 - (Y07660) B subunit of propionyl-CoA carboxylase [Mycobacterium tuberculosis]		carboxylase	1039
127	80074988 (18869, 18870)	Novel Protein sim. GBank gil 1877334 emb CAB07082 - (Z92771) bfrA [Mycobacterium tuberculosis]		carboxylase	1003, 1009, 1017, 1022, 1023, 1024, 1049
128	11698783 (8041, 8042)	Novel Protein sim. GBank gil 2266428 emb CAA74219 - (Y13917) yngE [Bacillus subtilis]		carboxylase	1031
129	79619609 (13081, 13082)	Novel Protein sim. GBank gil 25000015 sp P72158 PURK_PSEAE - PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE ATPASE SUBUNIT (AIR CARBOXYLASE) (AIRC)		carboxylase	1004
130	10153552 (20385, 20386)	Novel Protein sim. GBank gil 250015 sp P72158 PURK_PSEAE - PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE ATPASE SUBUNIT (AIR CARBOXYLASE) (AIRC)		carboxylase	1004

131	66489397 (17463, 17464)	Novel Protein sim. GBank gi 2627219 dbj BA23583] - (AB007000) propionyl- CoA carboxylase [Myxococcus xanthus]		carboxylase	1029
132	79850819 (10557, 10558)	Novel Protein sim. GBank gi 2708717 (AF038548) - pyruvate carboxylase [Corynebacterium glutamicum]		carboxylase	1044
133	95103879 (5671, 5672)	Novel Protein sim. GBank gi 4704652 gb AAD28194.1 (AF11360 - (AF113605) propionyl-CoA carboxylase complex B subunit [Streptomyces coelicolor])		carboxylase	1003, 1026
134	66556200 (13077, 13078)	Novel Protein sim. GBank gi 5420389 emb CAB46680.1 - (AJ243460) proteophosphoglycan [Leishmania major]		carboxylase	1023
135	21434541 (2473, 2474)	Novel Protein sim. GBank gi 585765 sp P32327 PYC2_YEAST - PYRUVATE CARBOXYLASE 2 (PYRUVIC CARBOXYLASE 2) (PCB_2)		carboxylase	1034
136	36512555 (22147, 22148)	Novel Protein sim. GBank gi 2266684 emb CAA67891] - (X99571) factor XII (Mus musculus)	Contains protein domain (PF00008) - EGF-like domain	cathepsin	1026
137	28991620 (8985, 8986)	Novel Protein sim. GBank gi 115718 sp P18242 CATD_MOUSE - CATHEPSIN D PRECURSOR	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	cathepsin	1044
138	28390646 (20713, 20714)	Novel Protein sim. GBank gi 115718 sp P18242 CATD_MOUSE - CATHEPSIN D PRECURSOR	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	cathepsin	1024, 1044
139	20729202 (8251, 8252)	Novel Protein sim. GBank gi 162858 (J00002) - preprochymosin a [Bos taurus]	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	cathepsin	1010
140	78778429 (10083, 10084)	Novel Protein sim. GBank gi 180352 (M20218) - coagulation factor XI [Homo sapiens]	Contains protein domain (PF00089) - Trypsin	cathepsin	1026
141	17659990 (7413, 7414)	Novel Protein sim. GBank gi 2266684 emb CAA67891] - (X99571) factor XII (Mus musculus)	Contains protein domain (PF00089) - Trypsin	cathepsin	1017

142	17681102 (18981, 18982)	Novel Protein sim. GBank gj 2266684 emb CAA67891 - (X99571) factor XII [Mus musculus]	Contains protein domain (PF000089) - Trypsin	cathepsin	1017
143	55869031 (16045, 16046)	Novel Protein sim. GBank gj 4 65317 dbj BAA37148 - (AB022426) FXII [Sus scrofa]	Contains protein domain (PF000089) - Trypsin	cathepsin	1009
144	30660979 (3297, 3298)	Novel Protein sim. GBank gj 547790 sp P36373 KLK7_RAT - GLANDULAR KALLIKREIN 7 PRECURSOR , SUBMANDIBULAR/RENAL (TISSUE KALLIKREIN) (RGK-7) (RSKG-7) (ESTERASE B) (PROTEINASE A)	Contains protein domain (PF000089) - Trypsin	cathepsin	1022
145	90936773 (1375, 1376)	Novel Protein sim. GBank gj 2746723 (AF038546) - cathepsin S precursor [Mus musculus]	Contains protein domain (PF00112) - Papain family cysteine protease	cathepsin	1020, 1024, 1044
146	20438750 (6429, 6430)	Novel Protein sim. GBank gj 2239204 emb CAB10121 - (Z97209) vacuolar carboxypeptidase [Schizosaccharomyces pombe]	Contains protein domain (PF00450) - Serine carboxypeptidase	cathepsin	1024
147	80503751 (21061, 21062)	Novel Protein sim. GBank gj 2499877 sp P70645 BLMH_RAT - BLEOMYCIN HYDROLASE (BLM HYDROLASE) (BMH)		cathepsin	1012, 1025
148	17898150 (5877, 5878)	Novel Protein sim. GBank gj 267476 sp Q01532 BLH1_YEAST - CYSTEINE PROTEINASE 1 (Y3) (BLEOMYCIN HYDROLASE) (BLM HYDROLASE)		cathepsin	1008
149	79566954 (16783, 16784)	Novel Protein sim. GBank gj 5305702 gb AAD41779_1 AF12686 - (AF126867) calpain-like protease [Mus musculus]		cathepsin	1001, 1008
150	8236940 (13219, 13220)	Novel Protein sim. GBank gj 1762152 (U51017) - kallistatin [Rattus norvegicus]	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	cathepsininhib	1022

151	8372988 (14285, 14286)	Novel Protein sim. GBank g 57233 emb CAA344071 - (X16358) SPI-2 serine protease inhibitor (AA 1-407) [Rattus norvegicus]	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	cathepsin inhib	1010
152	10250631 (6301, 6302)	Novel Protein sim. GBank g 165366 dbj BAA13578 - (D90915) chloride channel protein [Synechocystis sp.]		chloride_channel	1039
153	80236670 (6895, 6896)	Novel Protein sim. GBank g 4206711 (AF091390) - phosphotyramine precursor [Mus musculus]		chloride_channel	1000, 1004, 1008, 1017, 1022, 1023
154	19860823 (13483, 13484)	Novel Protein sim. GBank g 544380 sp P37020 GEF1 YEAST - GEF1 PROTEIN (VOLTAGE-GATED CHLORIDE CHANNEL) (CLC-Y1) (CLC-A)		chloride_channel	1022
155	77804342 (10571, 10572)	Novel Protein sim. GBank g 104615 pir TS20819 - collagen alpha 3(X) chain precursor - chicken	Contains protein domain (PF01391) - Collagen triple helix repeat (20 copies)	collagen	1009, 1011, 1038
156	30653868 (21363, 21364)	Novel Protein sim. GBank g 317200 emb CAA06511 - (AJ005396) collagen alpha 1 (XI) [Rattus norvegicus]	Contains protein domain (PF01410) - Fibrillar collagen C-terminal domain	collagen	1008
157	65495737 (5659, 5660)	Novel Protein sim. GBank g 1076057 pir PC4028 - dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Saccharopolyspora erythraea (fragment)		collagen	1023
158	79817511 (1807, 1808)	Novel Protein sim. GBank g 1150834 (U42471) - Wiscott-Aldrich Syndrome protein homolog [Mus musculus]		collagen	1044, 1054
159	37796258 (1121, 1122)	Novel Protein sim. GBank g 34398 sp P15040 SEC_B_ECOLI - PROTEIN- EXPORT PROTEIN SEC_B		collagen	1029
160	65860855 (11265, 11266)	Novel Protein sim. GBank g 1345652 sp P15989 CA36_CHICK - COLLAGEN ALPHA 3(V) CHAIN PRECURSOR		collagen	1022, 1038
161	79482900 (16169, 16170)	Novel Protein sim. GBank g 1493807 (U35068) - PxrR [Pseudomonas aeruginosa]		collagen	1006

162	86473821 (22817, 22818)	Novel Protein sim. GBank gi 2098284 pdb 1SRR A - Chain A, Crystal Structure Of A Phosphatase Resistant Mutant Of Sporulation Response Regulator Spo0 From <i>Bacillus Subtilis</i>		collagen	1029
163	69930092 (5617, 5618)	Novel Protein sim. GBank gi 2493324 sp Q28107 FA5_BOVIN - COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR)		collagen	1053
164	99435210 (233, 234)	Novel Protein sim. GBank gi 323091 pir A45638 - immunodominant microneme protein Etp100 - <i>Eimeria tenella</i>		collagen	1010, 1016, 1024, 1051, 1058
165	79560182 (13263, 13264)	Novel Protein sim. GBank gi 4582324 emb CAB40275.1 - (AL031782) dj708F5.1 (PUTATIVE novel Collagen alpha 1 LIKE protein) (<i>Homo sapiens</i>)		collagen	1038
166	24128147 (7013, 7014)	Novel Protein sim. GBank gi 517464 ref NP_006106.1 pPRAM - preferentially expressed antigen of melanoma		collagen	1003
167	29228285 (18405, 18406)			collagen	1030
168	78770688 (19125, 19126)			collagen	1001, 1008
169	85526555 (21579, 21580)			collagen	1001, 1003, 1008, 1016, 1017, 1023, 1024, 1025, 1029, 1034, 1038, 1044, 1049
170	80226969 (18787, 18788)	Novel Protein sim. GBank gi 2144487 pir HPRT - haptoglobin precursor - rat	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	1003, 1010, 1039
171	27845657 (13391, 13392)	Novel Protein sim. GBank gi 2493792 sp Q63514 C4BP_RAT - C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	1010

172 30640031 (18421, 18422)	Novel Protein sim. GBank gi 2493792 sp Q63514 C4BP_RAT - C4B- BINDING PROTEIN ALPHA CHAIN PRECURSOR	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	1022
173 78368514 (22409, 22410)	Novel Protein sim. GBank gi 543375 pir JC2054 - precursor -rat	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	1003
174 95010147 (15783, 15784)	Novel Protein sim. GBank gi 285190 pir A42210 - alpha-1-macroglobulin precursor - rat	Contains protein domain (PF00207) - Alpha-2- macroglobulin family	complement	1008
175 11309122 (245, 246)	Novel Protein sim. GBank gi 112883 sp P14046 A113_RAT - ALPHA-1- INHIBITOR III PRECURSOR	Contains protein domain (PF01835) - Alpha-2- macroglobulin family N- terminal region	complement	1010
176 25152279 (12809, 12810)	Novel Protein sim. GBank gi 1256828 U52948 - complement component C9 precursor [Rattus norvegicus]		complement	1034
177 30578467 (18561, 18562)	Novel Protein sim. GBank gi 3915642 sp P06681 CO2_HUMAN - COMPLEMENT C2 PRECURSOR (C3/C5 CONVERTASE)		complement	1006, 1026, 1044
178 23332011 (10511, 10512), 80052954 (12063, 12064)	Novel Protein sim. GBank gi 2130120 pir B57742 - cyclin lb - maize	Contains protein domain (PF00134) - Cyclin	cyclin	1053
179	Novel Protein sim. GBank gi 731808 sp P40186 YFO_YEAST - HYPOTHEtical 32.0 KD PROTEIN IN GPP1- SYG1 INTERGENIC REGION		cyclin	1017, 1022, 1024, 1031, 1039
180 79844358 (9089, 9090)	Novel Protein sim. GBank gi 1117170 sp P24464 CP48_RAT - CYTOCHROME P450 4A8 (CYP4A8) (P450- KP1) (P450-PP1)	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1017, 1039
181 78917934 (5265, 5266)	Novel Protein sim. GBank gi 117237 sp P24470 CPCN_RAT - CYTOCHROME P450 2C23 (CYP2C23)	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1017, 1026

182 30663025 (8923, 8924)	Novel Protein sim. GBank gil117254 ispIP24461 CPG1_RABIT - CYTOCHROME P450 2G1 (CYP1G1) (P450- NMB) (OLFACTIVE)	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1022
183 79332599 (2525, 2526)	Novel Protein sim. GBank gil1352184 ispIP20817 CP43_RAT - CYTOCHROME P450 4A3 (CYP1VA3) (LAURIC ACID OMEGA-HYDROXYLASE) (P450-LA- OMEGA 3)	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1008, 1024, 1033, 1034, 1044
184 856812055 (20661, 20662)	Novel Protein sim. GBank gil203695 (M18363) - cytochrome P450 [Rattus norvegicus]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1026
185 80586214 (19753, 19754)	Novel Protein sim. GBank gil2117398 pir A46588 - cytochrome P450 arachidonic acid epoxygenase - rat	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1022, 1058
186 80054009 (20993, 20994)	Novel Protein sim. GBank gil2344865 emb CAA72345 - (Y11638) cytochrom3 P450 [Mus musculus]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1022, 1025
187 78971513 (20737, 20738)	Novel Protein sim. GBank gil3461849 (AC005315) - putative cytochrome P450 [Arabidopsis thaliana]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1008
188 79191241 (2825, 2826)	Novel Protein sim. GBank gil499858 (U09734) - testosterone 6-beta-hydroxylase [Rattus norvegicus]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1017, 1022, 1024, 1053
189 28814552 (10103, 10104)	Novel Protein sim. GBank gil499858 (U09734) - testosterone 6-beta-hydroxylase [Rattus norvegicus]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1044
190 25126609 (3329, 3330)	Novel Protein sim. GBank gil55054 emb CAA57425 - (X81831) cytochrome P450 [Zea mays]	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1008
191 17295741 (4963, 4964)	Novel Protein sim. GBank gil7979741 pir C37842 - cytochrome P450 hydroxylase homolog - Anabaena sp. (PCC 7120)	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1008
192 27832970 (455, 456)	Novel Protein sim. GBank gil92129 pir A36122 - cytochrome P450 2C13 - rat	Contains protein domain (PF00067) - Cytochrome P450	cyto450	1006

193	11077811 (4783, 4784)	Novel Protein sim. GBank gi 1171579 emb CAA64635I - (X95342) cytochrome P450 [Nicotiana tabacum]		cyto450	1010
194	79625816 (10731, 10732)	Novel Protein sim. GBank gi 1352185 sp P49602 CP51_UTMMA - CYTOCHROME P450_51 (CYPL1) (P450-L1A1) (STEROL 14-ALPHA DEMETHYLASE) (LANSOSTEROL 14-ALPHA DEMETHYLASE) (P450-14DM)		cyto450	1022
195	80592414 (21177, 21178)	Novel Protein sim. GBank gi 2143679 pir IB56755 - cytochrome P450 2C24, splice form p9 - rat (fragment)		cyto450	1000, 1003, 1004, 1006, 1008, 1009, 1022, 1023, 1024, 1029, 1034, 1039, 1058
196	65508850 (16151, 16152)	Novel Protein sim. GBank gi 2995384 emb CAA06156I - (AJ004810) cytochrome P450 monooxygenase [Zea mays]		cyto450	1004, 1009
197	16329769 (16105, 16106)	Novel Protein sim. GBank gi 4503213 ref NP_000776.1 CYP2 - cytochrome P450, subfamily XXVIIIB (25-hydroxyvitamin D-1- alpha-hydroxylase), polypeptide 1		cyto450	1034
198	80504735 (11071, 11072)			cyto450	1012
199	78753909 (12695, 12696)	Novel Protein sim. GBank gi 3929343 sp O31215 CYB_CHRVI - CYTOCHROME B	Contains protein domain (PF00032) - Cytochrome b(C terminal)/B6/petD	cytochrome	1003
200	65692479 (13805, 13806)		Contains protein domain (PF00034) - Cytochrome c	cytochrome	1009
201	78678334 (425, 426)	Novel Protein sim. GBank gi 3915154 sp Q12612 TR14_FUSSP - TRICHODIENE OXYGENASE (CYTOCHROME P450_58)	Contains protein domain (PF00067) - Cytochrome P450	cytochrome	1017

202	39517846 (21927, 21928)	Novel Protein sim. GBank g 1071819 pir B54759 - cytochrome ba(3) chain I - Paracoccus denitrificans	- Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	cytochrome	1006
203	78259702 (21693, 21694)	Novel Protein sim. GBank g 4539224 emb CAB39882.1 - (AL049497) putative cytochrome c oxidase subunit I [Streptomyces coelicolor]	- Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	cytochrome	1026
204	9548161 (18807, 18808)	Novel Protein sim. GBank g 14932 sp P11056 BFR_ECOLI - BACTERIOFERRITIN (BFR) (CYTOCHROME B-1) (CYTOCHROME B-557)	- Contains protein domain (PF01334) - Bacterioferritin	cytochrome	1001
205	78693128 (9487, 9488)	Novel Protein sim. GBank g 168839 sp P45037 CCMF_HAEIN - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF	-	cytochrome	1008, 1054
206	10246324 (17651, 17552)	Novel Protein sim. GBank g 169161 sp P00106 CYC4_PSEAE - CYTOCHROME C4	-	cytochrome	1023
207	19534127 (15257, 15258)	Novel Protein sim. GBank g 1705703 sp P52225 CCMF_PSEFL - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK	-	cytochrome	1010
208	13085830 (7847, 7848)	Novel Protein sim. GBank g 1706225 sp P51131 CYB_BRAJA - CYTOCHROME B/C1 PRECURSOR	-	cytochrome	1027
209	12992395 (8403, 8404)	Novel Protein sim. GBank g 1722914 sp Q10388 Y007_MYCTU - HYPOTHETICAL 61.0 KD PROTEIN CY190.07	-	cytochrome	1024
210	80052571 (22491, 22492)	Novel Protein sim. GBank g 466191 sp P35160 RESA_BACSU - RESA PROTEIN	-	cytochrome	1000, 1010, 1026
211	1324611 (16627, 16628)	Novel Protein sim. GBank g 116041 sp P19079 CDD_BACSU - CYTIDINE DEAMINASE (CYTIDINE AMINOHYDROLASE) (CDA)	- Contains protein domain (PF00383) - Cytidine and deoxycytidine deaminase zinc-binding region	deaminase	1027

212	29444838 (6193, 8194)	Novel Protein sim. GBank g 2956780 emb CAA17893_1 - (AL022103) putative deoxycytidylate deaminase [Schizosaccharomyces pombe]	Contains protein domain (PF00383) - Cytidine and deoxycytidylate deaminase zinc-binding region	deaminase	1026
213	79910020 (17693, 17694)	Novel Protein sim. GBank g 2960176 emb CAA18074_1 - (AL022121) hypothetical protein Rv3752c [Mycobacterium tuberculosis]	Contains protein domain (PF00383) - Cytidine and deoxycytidylate deaminase zinc-binding region	deaminase	1004, 1008, 1031, 1038
214	10858749 (19165, 19166)	Novel Protein sim. GBank g 1166628 U46753 - coded for by C. elegans cDNA cm217; coded for by C. elegans cDNA yk33e12.5; coded for by C. elegans cDNA cm508; coded for by C. elegans cDNA cm08d9; coded for by C. elegans cDNA yk42f12.5; coded for by C. elegans cDNA yk6a8.5; code...		deaminase	1031
215	65506178 (14541, 14542)	Novel Protein sim. GBank g 1177607 emb CAA63219 - (X92485) pva1 [Plasmodium vivax]		deaminase	1016
216	20289268 (263, 264)	Novel Protein sim. GBank g 1723726 sp P53294 YG3X_YEAST - HYPOTHETICAL 46.8 KD PROTEIN IN CLC1-PDS2 INTERGENIC REGION		deaminase	1010
217	80481756 (9253, 9254)	Novel Protein sim. GBank g 2314571 sp P302971 A1D_PSES0 - 1-AMINOACYLCLOPROPANE-1-CARBOXYLATE DEAMINASE (ACC DEAMINASE)		deaminase	1012
218	79111827 (9559, 9560)	Novel Protein sim. GBank g 2501674 sp Q12362 RIB2_YEAST - DRAP DEAMINASE		deaminase	1024, 1044
219	78781506 (18543, 18544)	Novel Protein sim. GBank g 3915436 sp Q12069 YD36_YEAST - HYPOTHETICAL 53.4 KD PROTEIN IN PRP9-NAT1 INTERGENIC REGION		deaminase	1022

220	78782067 (19367, 19368)	Novel Protein sim. GBank gi 3915436 sp Q12069 YD36_YEAST - HYPOTHETICAL 53.4 KD PROTEIN IN PRP9 - NAT1 INTERGENIC REGION		deaminase	1022
221	10886316 (14007, 14008)	Novel Protein sim. GBank gi 417043 sp P32064 GCVA_ECOLI - GLYCINE CLEAVAGE SYSTEM TRANSCRIPTIONAL ACTIVATOR (GCV OPERON ACTIVATOR)		deaminase	1022
222	29691266 (14421, 14422)	Novel Protein sim. GBank gi 446602 sp CAB37622 - (AL035581) putative threonine dehydratase precursor [Schizosaccharomyces pombe]		deaminase	1024
223	85800967 (21413, 21414)	Novel Protein sim. GBank gi 483532 sp CAB42949_1 - (AL049863) putative adenosine deaminase [Streptomyces coelicolor]		deaminase	1025, 1026, 1038
224	17876454 (6033, 6034)	Novel Protein sim. GBank gi 728838 sp P39195 ALU8_HUMAN - !!! ALU SUBFAMILY SX WARNING ENTRY !!!		deaminase	1024
225	79821726 (13935, 13936)	Novel Protein sim. GBank gi 120538 sp Q00236 NUIC_PLEBO - NADH- PLASTOQUINONE OXIDOREDUCTASE SUBUNIT 1 (FRXB PROTEIN)	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron- sulfur cluster binding domains.	dehydrogenase	1003, 1024, 1026, 1034
226	27978888 (16887, 16888)	Novel Protein sim. GBank gi 3273346 dbj BA31216 - (AB015757) fumarate reductase iron-sulphur protein subunit [Rhodoflexax fermentans]	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron- sulfur cluster binding domains.	dehydrogenase	1022
227	8504263 (8013, 8014)	Novel Protein sim. GBank gi 504226 sp CAB44523.1 - (AL078618) nuo, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron- sulfur cluster binding domains.	dehydrogenase	1010
228	21425187 (10297, 10298)	Novel Protein sim. GBank gi 126053 sp P14561 DHP_BACPS - L- LACTATE DEHYDROGENASE P	Contains protein domain (PF00056) - lactate/malate dehydrogenase	dehydrogenase	1022

229	20457067 (895, 896)	Novel Protein sim. GBank gi 2836200 (AF023168) - L-lactate dehydrogenase [Styela pilicata]	Contains protein domain (PF00056) - lactate dehydrogenase	dehydrogenase	1010
230	30187977 (9057, 9058)	Novel Protein sim. GBank gi 547837 sp P16115 LDH_THEME - L-LACTATE DEHYDROGENASE	Contains protein domain (PF00056) - lactate/malate dehydrogenase	dehydrogenase	1009
231	11687065 (9869, 9870)	Novel Protein sim. GBank gi 547837 sp P16115 LDH_THEME - L-LACTATE DEHYDROGENASE	Contains protein domain (PF00056) - lactate/malate dehydrogenase	dehydrogenase	1006
232	34394280 (20873, 20874)	Novel Protein sim. GBank gi 3024243 sp P95174 NUOJ_MYCTU - NADH DEHYDROGENASE	Contains protein domain (PF00146) - NADH dehydrogenases	dehydrogenase	1026
233	9844371 (2383, 2384)	Novel Protein sim. GBank gi 3861323 emb CAA15222 - (AJ235273) NADH DEHYDROGENASE I CHAIN H (nuoH) [Rickettsia prowazekii]	Contains protein domain (PF00146) - NADH dehydrogenases	dehydrogenase	1023
234	80226363 (21519, 21520)	Novel Protein sim. GBank gi 66132 pif QXMMS1M - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - mouse mitochondrion (SGC1)	Contains protein domain (PF00146) - NADH dehydrogenases	dehydrogenase	1003, 1004, 1008, 1009, 1016, 1017, 1022, 1023, 1038, 1039, 1044
235	11076591 (21485, 21486)	Novel Protein sim. GBank gi 1182011 emb CAA60637 - (X87172) NAD+-isocitrate dehydrogenase, alpha subunit [Macaca fascicularis]	Contains protein domain (PF00180) - Isocitrate and isopropylmalate dehydrogenases	dehydrogenase	1010
236	78674526 (6535, 6536)	Novel Protein sim. GBank gi 124160 sp P21954 IDHP YEAST - ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR (OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDDH) (IDP)	Contains protein domain (PF00180) - Isocitrate and isopropylmalate dehydrogenases	dehydrogenase	1003, 1009, 1017, 1022

237	10885283 (21185, 21196)	Novel Protein sim. GBank g 1261971 sp P24404 ILEU3_AGRTU - 3-ISOPROPYLALATE DEHYDROGENASE (BETA) (3-IPM-DH)	Contains protein domain (PF00180) - isocitrate and isopropylmalate dehydrogenases	dehydrogenase	1024
238	79574361 (7043, 7044)	Novel Protein sim. GBank g 13355670 emb CAA19970 - (AL031124) - 3-isopropylmalate dehydrogenase [Streptomyces coelicolor]	Contains protein domain (PF00180) - isocitrate and isopropylmalate dehydrogenases	dehydrogenase	1031
239	80250117 (3313, 3314)	Novel Protein sim. GBank g 462384 sp P33198 IDHP_PIG - ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR (OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+)-SPECIFIC (ICDH) (IDP) (ICD-M)	Contains protein domain (PF00180) - isocitrate and isopropylmalate dehydrogenases	dehydrogenase	1022
240	80503433 (20213, 20214)	Novel Protein sim. GBank g 3219722 (AF0668740) - dihydrolipoamide succinyltransferase; E2 [Pseudomonas putida]	Contains protein domain (PF00198) - 2-oxo acid dehydrogenases acyltransferase (catalytic domain)	dehydrogenase	1012
241	17659515 (3023, 3024)	Novel Protein sim. GBank g 171594 (M10590) - glutamate dehydrogenase [Saccharomyces cerevisiae]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	1023
242	80230771 (953, 954)	Novel Protein sim. GBank g 322228 p t (S322227 - glutamate dehydrogenase (NADP+)) (EC 1.4.1.4) - Corynebacterium glutamicum	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	1022, 1023, 1024, 1034
243	79831058 (17321, 17322)	Novel Protein sim. GBank g 4239787 emb CAA75437 - (Y15166) NADP-glutamate dehydrogenase [Pseudomonas aeruginosa]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	1023

244 4370)	78748880 (4369, Novel Protein sim. GBank gi 461927 sp P33327 DHE2 YEAST - NAD- SPECIFIC GLUTAMATE DEHYDROGENASE (NAD-GDH)	Contains protein domain (PF00208) - Glutamate/Phenylal- anine/Valine dehydrogenase	dehydrogenase	1003, 1008, 1017, 1022
245 18792)	79611609 (18791, Novel Protein sim. GBank gi 4981555 gb AD36092.1 AE00176 - (AE001763) glutamate dehydrogenase [Thermotoga maritima]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylal- anine/Valine dehydrogenase	dehydrogenase	1006
246 10180)	79377592 (10179, Novel Protein sim. GBank gi 3378650 emb CAA66205 - (X97606) abscisic acid activated [Medicago sativa]	Contains protein domain (PF00248) - Aldo/keto reductase family	dehydrogenase	1004
247 9994)	80071700 (9993, Novel Protein sim. GBank gi 4585838 emb CAB40932.1 - (ALD49630) putative NADH dehydrogenase [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory- chain NADH dehydrogenase, 49 Kd subunit	dehydrogenase	1001, 1022, 1023, 1026
248 2244)	66727102 (2243, Novel Protein sim. GBank gi 5042274 emb CAB44528.1 - (AL078618) nuoD, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory- chain NADH dehydrogenase, 49 Kd subunit	dehydrogenase	1009, 1024
249 1870)	56649616 (1869, Novel Protein sim. GBank gi 128764 sp P05508 NU4M RAT - NADH- UBIQUINONE OXIDOREDUCTASE CHAIN 4	Contains protein domain (PF00361) - NADH- Ubiquinone/plastoquinone (complex I), various chains	dehydrogenase	1008, 1016, 1038, 1044
250 9072)	10126621 (9071, Novel Protein sim. GBank gi 5042264 emb CAB44518.1 - (AL078618) nuoN, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00361) - NADH- Ubiquinone/plastoquinone (complex I), various chains	dehydrogenase	1038
251 7192)	80079351 (7191, Novel Protein sim. GBank gi 5042266 emb CAB44520.1 - (AL078618) nuoN, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00361) - NADH- Ubiquinone/plastoquinone (complex I), various chains	dehydrogenase	1034

252	21639409 (1017, 1018)	Novel Protein sim. GBank gil541157 pir S38321 - NADH dehydrogenase I chain nuol • Escherichia coli	Contains protein domain (PF00361) - NADH-Ubiquinone/plastoquinone (complex I), various chains (contains protein domain (PF00361) - NADH-Ubiquinone/plastoquinone (complex I), various chains (contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	dehydrogenase	1044
253	78669607 (16915, 16916)	Novel Protein sim. GBank gil755859 (U21985) - NADH dehydrogenase F [Zea mays]	Contains protein domain (PF00361) - NADH-Ubiquinone/plastoquinone (complex I), various chains (contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	dehydrogenase	1009
254	80063153 (437, 438)	Novel Protein sim. GBank gil2494086 sp P75913 YCDW_ECOLI - PUTATIVE 2-HYDROXYACID DEHYDROGENASE IN PHOH-CSGG INTERGENIC REGION	Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	dehydrogenase	1024, 1034
255	10867193 (17609, 17610)	Novel Protein sim. GBank gil325703 dbj BA29686 I - (AP000002) 376aa long hypothetical dehydrogenase [Pyrococcus horikoshii]	Contains protein domain (PF00389) - D-isomer specific 2-hydroxyacid dehydrogenases	dehydrogenase	1006
256	946811793 (11863, 11864)	Novel Protein sim. GBank gil100506 pir S17455 - Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - Flaveria trinervia (fragment)	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1031
257	28387444 (3603, 3604)	Novel Protein sim. GBank gil1170871 sp PA3279 MAOC_ORYSA - MALATE OXIDOREDUCTASE, CHLOROPLAST PRECURSOR (MALIC ENZYME) (ME) (NADP-DEPENDENT MALIC ENZYME) (NADP-ME)	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1017
258	80059740 (4575, 4576)	Novel Protein sim. GBank gil2293209 (AF008220) - putative malolactic enzyme [Bacillus subtilis]	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1024
259	32691341 (20477, 20478)	Novel Protein sim. GBank gil2293209 (AF008220) - putative malolactic enzyme [Bacillus subtilis]	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1029
260	30653334 (18715, 18716)	Novel Protein sim. GBank gil3047125 (AF058919) - No definition line found [Arabidopsis thaliana]	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1022

261	28811526 (167, 168)	Novel Protein sim. GBank g 4505145 ref NP_002387.1 pME2 - malic enzyme 2, mitochondrial	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1004, 1009, 1024
262	32358210 (10595, 10596)	Novel Protein sim. GBank g 549664 sp P36099 YKDO_YEAST - HYPOTHETICAL 23.0 KD PROTEIN IN IXR1-TFA1 INTERGENIC REGION	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	1016
263	78925330 (11703, 11704)	Novel Protein sim. GBank g 123755 sp P12380 HUTC_KLEAE - HISTIDINE UTILIZATION REPRESSOR	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	dehydrogenase	1039
264	8361988 (16665, 16666)	Novel Protein sim. GBank g 1346565 sp P49309 MOCR_RHIME - PROBABLE RHIZOPINE CATABOLISM REGULATORY PROTEIN MOCR	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	dehydrogenase	1034
265	27971818 (15893, 15894)	Novel Protein sim. GBank g 2129478 pir S51939 - chitinase (EC 3.2.1.14) precursor - beet	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	dehydrogenase	1034
266	80217174 (15305, 15306)	Novel Protein sim. GBank g 112842 sp P14332 6PGD_PIG - 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING	Contains protein domain (PF00393) - 6-phosphogluconate dehydrogenases	dehydrogenase	1006, 1017, 1022
267	78890715 (14173, 14174)	Novel Protein sim. GBank g 1781203 emb CAB06110 - (233859) gnd [Mycobacterium tuberculosis]	Contains protein domain (PF00393) - 6-phosphogluconate dehydrogenases	dehydrogenase	1003
268	17299964 (19373, 19374)	Novel Protein sim. GBank g 165945 pir DEECGC - phosphogluconate dehydrogenase (decarboxylating) (EC 1.1.1.44) - Escherichia coli	Contains protein domain (PF00393) - 6-phosphogluconate dehydrogenases	dehydrogenase	1008
269	9655278 (73, 74)	Novel Protein sim. GBank g 111334 pir B30605 - acyl-CoA dehydrogenase (EC 1.3.99.3) precursor, short-chain-specific - rat	Contains protein domain (PF0441) - Acyl-CoA dehydrogenase	dehydrogenase	1034

270	78910477 (20229, 20230)	Novel Protein sim. GBank g 113018 sp P08503 ACDM_RAT - ACYL-COA DEHYDROGENASE, MEDIUM-CHAIN SPECIFIC PRECURSOR (MCAD)	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1008, 1009, 1017, 1039, 1044
271	10195284 (19277, 19278)	Novel Protein sim. GBank g 1877279 emb CAE07030 - (Z92770) fadE2 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1003
272	17841345 (2517, 2518)	Novel Protein sim. GBank g 3261533 emb CAA17469 - (AL021943) fadE8 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1001
273	21425904 (661, 662)	Novel Protein sim. GBank g 450479 ref NP_002216.1 pIVD - isovaleryl Coenzyme A dehydrogenase	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1022
274	65708031 (7121, 7122)	Novel Protein sim. GBank g 4582976 gb AAD24879.1 AF08192 - (AF081920) putative acyl-CoA dehydrogenase; PfTE [Pseudomonas fluorescens]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1023
275	9672090 (11177, 11178)	Novel Protein sim. GBank g 543789 sp P33224 AIDB_ECOLI - AIDB PROTEIN	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	1034
276	20710955 (19401, 19402)	Novel Protein sim. GBank g 1732040 U55016 - alcohol dehydrogenase [Actinobacillus pleuropneumoniae]	Contains protein domain (PF00465) - Iron-containing alcohol dehydrogenases	1006
277	19872109 (14599, 14600)	Novel Protein sim. GBank g 20735 sp P21907 G6PD_ZYMMO - GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (G6PD)	Contains protein domain (PF00479) - Glucose-6-phosphate dehydrogenase	1022
278	80059873 (467, 468)	Novel Protein sim. GBank g 3334792 emb CAA19940 - (AL031107) glucose-6-phosphate 1-dehydrogenase [Streptomyces coelicolor]	Contains protein domain (PF00479) - Glucose-6-phosphate dehydrogenase	1003, 1010, 1024, 1031
279	33745838 (20145, 20146)	Novel Protein sim. GBank g 4210900 gb AAD12043.1 - (AF045609) Zwf [Sinorhizobium meliloti]	Contains protein domain (PF00479) - Glucose-6-phosphate dehydrogenase	1026

280	79764146 (5165, 5166)	Novel Protein sim. GBank gil495200 (U09637) - NADH dehydrogenase subunit 3 [Mus musculus]	Contains protein domain (PF00507) - NADH-ubiquinone/plastoquinone oxidoreductase, chain 3	dehydrogenase	1023
281	80218245 (5745, 5746)	Novel Protein sim. GBank gil486964 pir S36713 - CCR4 protein - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00560) - Leucine Rich Repeat	dehydrogenase	1001, 1004, 1006, 1022, 1034
282	11392317 (15007, 15008)	Novel Protein sim. GBank gil2497360 spl Q50715IMDH_MYCTU - INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP DEHYDROGENASE) (IMPDH) (IMPDH)	Contains protein domain (PF00571) - CBS domain	dehydrogenase	1004
283	10075692 (19185, 19186)	Novel Protein sim. GBank gil1790870 (U32622) - toluenesulfonate zinc-independent alcohol dehydrogenase [Comamonas testosteron]	Contains protein domain (PF00678) - Short chain dehydrogenase/reductase C-terminus	dehydrogenase	1017
284	95010836 (1075, 1076)	Novel Protein sim. GBank gil2633748 emb CAB13250 - (Z99111) similar to glucose 1'-dehydrogenase [Bacillus subtilis]	Contains protein domain (PF00678) - Short chain dehydrogenase/reductase C-terminus	dehydrogenase	1017
285	47656792 (10983, 10984)	Novel Protein sim. GBank gil23266687 emb CAB10998 - (Z98271) methylenetetrahydrofolate dehydrogenase [Mycobacterium leprae]	Contains protein domain (PF00763) - Tetrahydrofolate dehydrogenase/cyclohydrolase	dehydrogenase	1029
286	37021456 (5707, 5708)	Novel Protein sim. GBank gil1666123 emb CAB05041 - (Z82098) hypothetical protein Rv3537 [Mycobacterium tuberculosis]	Contains protein domain (PF00890) - FAD binding domain	dehydrogenase	1025
287	79796075 (3207, 3208)	Novel Protein sim. GBank gil66666 pir DEECSF - succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Escherichia coli	Contains protein domain (PF00890) - FAD binding domain	dehydrogenase	1044
288	79853922 (9681, 9682)	Novel Protein sim. GBank gil66666 pir DEECSF - succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Escherichia coli	Contains protein domain (PF00890) - FAD binding domain	dehydrogenase	1044

289	80066937 (5299, 5300)	Novel Protein sim. GBank gj 216556 dbj BAA02174 - (D12651) glucose dehydrogenase [Escherichia coli]	Contains protein domain (PF01011) - PQQ enzyme repeat	dehydrogenase	1024
290	85802878 (3151, 3152)	Novel Protein sim. GBank gj 3183438 sp P77774 YFGL_ECOLI - HYPOTHETICAL 4.1.9 KD PROTEIN IN XSEA-HISS INTERGENIC REGION	Contains protein domain (PF01011) - PQQ enzyme repeat	dehydrogenase	1054
291	79552944 (6951, 6952)	Novel Protein sim. GBank gj 1877346 emb CA807073 - (Z92771) hypothetical protein Rv3291c [Mycobacterium tuberculosis]	Contains protein domain (PF01037) - AsnC family	dehydrogenase	1025, 1044
292	86467342 (6999, 7000)	Novel Protein sim. GBank gj 3342734 (AC005329) - NUKM_HUMAN, partial CDS; COMPILE; Cl-20KD; PSST SUBUNIT [Homo sapiens]	Contains protein domain (PF01058) - NADH ubiquinone oxidoreductase, 20 Kd subunit	dehydrogenase	1029
293	80502307 (10955, 10956)	Novel Protein sim. GBank gj 2497380 sp Q50715 IMDH_MYCTU - INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP DEHYDROGENASE) (IMPDH) (IMPD)	Contains protein domain (PF01574) - IMP dehydrogenase / GMP reductase N terminus	dehydrogenase	1012
294	79840419 (10967, 10968)	Novel Protein sim. GBank gj 2495499 sp P77165 YAGT_ECOLI - HYPOTHETICAL 24.3 KD PROTEIN IN INTF-EAEH INTERGENIC REGION	Contains protein domain (PF01799) - [2Fe-2S] binding domain	dehydrogenase	1008
295	86688076 (5465, 5466)	Novel Protein sim. GBank gj 56689912 emb CAE52075.1 - (AL109732) putative mutase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01817) - Chorismate mutase	dehydrogenase	1006, 1008, 1010, 1012, 1017, 1019, 1022, 1025, 1026, 1029, 1038, 1039, 1053
296	80074011 (21089, 21090)	Novel Protein sim. GBank gj 1071808 pir S50576 - probable aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) YER073w - yeast (Saccharomyces cerevisiae)		dehydrogenase	1008, 1022, 1023, 1038

297	14998014 (14371, 14372)	Novel Protein sim. GBank gj 1083428 pir S54876 - NAD(P)+ transhydrogenase (B-specific) (EC 1.6.1.1) precursor - mouse		dehydrogenase	1024
298	87453009 (12295, 12296)	Novel Protein sim. GBank gj 1084349 pir S53503 - aldehyde dehydrogenase homolog btg-26 - rape		dehydrogenase	1025
299	79778373 (14419, 14420)	Novel Protein sim. GBank gi 1122429 emb CAA55142I - (X78346) proline dehydrogenase [Rhodobacter capsulatus]		dehydrogenase	1017
300	65474140 (7049, 7050)	Novel Protein sim. GBank gj 1166604 (U46673) - Similar to aldehyde dehydrogenase [Caenorhabditis elegans]		dehydrogenase	1038
301	16730108 (20847, 20848)	Novel Protein sim. GBank gi 1168352 spi P41681 ADH2_PERMA - ALCOHOL DEHYDROGENASE 2		dehydrogenase	1017
302	80059940 (19643, 19644)	Novel Protein sim. GBank gi 1169286 spi P42757 DHAB_ATRHO - BETAINE- ALDEHYDE DEHYDROGENASE PRECURSOR (BADH)		dehydrogenase	1003, 1010
303	87467703 (7875, 7876)	Novel Protein sim. GBank gi 1169603 spi P46154 FADH_PSEPU - GLUTATHIONE-INDEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH) (FALDH)		dehydrogenase	1018
304	79914093 (20509, 20510)	Novel Protein sim. GBank gi 1169793 spi P46713 G3P_MYCLE - GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)		dehydrogenase	1038
305	78459708 (12449, 12450)	Novel Protein sim. GBank gi 1169945 spi P43799 GLPA_HAEIN - ANAOEROBIC GLYCEROL-3-PHOSPHATE DEHYDROGENASE SUBUNIT A (G-3-P DEHYDROGENASE)		dehydrogenase	1024, 1026

306	43153908 (18471, 18472)	Novel Protein sim. GBank gi 1170354 isp P42269 HPCC_ECOLI - 5-CARBOXYMETHYL-2-HYDROXYMUCONATE SEMIALDEHYDE DEHYDROGENASE (CHMS) DEHYDROGENASE		dehydrogenase	1000
307	24124818 (10591, 10592)	Novel Protein sim. GBank gi 1171071 isp P44605 MURB_HAEIN - UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE REDUCTASE (UDP-N-ACETYLGLUCOSAMINE DEHYDROGENASE)		dehydrogenase	1044
308	27845642 (11167, 11168)	Novel Protein sim. GBank gi 1171948 isp P41816 OYE3_YEAST - NADPH DEHYDROGENASE 3 (OLD YELLOW ENZYME 3)		dehydrogenase	1008, 1010
309	78678687 (21551, 21552)	Novel Protein sim. GBank gi 1175116 isp P43939 Y094_HAEIN - HYPOTHETICAL PROTEIN HI0094		dehydrogenase	1003
310	57108030 (2885, 2886)	Novel Protein sim. GBank gi 117528 isp P14755 ICRYL_RABIT - LAMBDA-CRYSTALLIN		dehydrogenase	1022
311	66373023 (8447, 8448)	Novel Protein sim. GBank gi 118227 isp P29011 DADA_ECOLI - D-AMINO ACID DEHYDROGENASE SMALL SUBUNIT		dehydrogenase	1038
312	27972348 (5585, 5586)	Novel Protein sim. GBank gi 118563 isp P18172 DHGL_DROPS - GLUCOSE DEHYDROGENASE (ACCEPTOR) PRECURSOR		dehydrogenase	1022
313	80187718 (6953, 6954)	Novel Protein sim. GBank gi 118619 isp P21801 DHSB_YEAST - SUCCINATE DEHYDROGENASE (UBIQUINONE) IRON-SULFUR PROTEIN PRECURSOR (IP)		dehydrogenase	1058
314	20606685 (14697, 14698)	Novel Protein sim. GBank gi 120712 isp P00361 G3P_THEAQ - GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)		dehydrogenase	1004

315	86685293 (12499, 12500)	Novel Protein sim. GBank gil120777 sp P25526 GABD_ECOLI - SUCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (SSDH)		dehydrogenase	1022, 1029, 1030
316	10213436 (20097, 20098)	Novel Protein sim. GBank gil1230584 (U47057) - nucleotide sugar dehydrogenase [Vibrio cholerae O139]		dehydrogenase	1024
317	82050554 (1661, 1662)	Novel Protein sim. GBank gil129036 sp P20707 ODO1_AZQVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)		dehydrogenase	1004, 1006, 1010, 1012, 1016, 1017, 1022
318	77520505 (1993, 1994)	Novel Protein sim. GBank gil130017 sp P23238 PHBB_ZOORA - ACETOACETYL-COA REDUCTASE		dehydrogenase	1049
319	78741921 (8941, 8942)	Novel Protein sim. GBank gil135188 sp P07246 ADH3_YEAST - ALCOHOL DEHYDROGENASE III PRECURSOR		dehydrogenase	1026, 1044
320	79822_50 (11839, 11840)	Novel Protein sim. GBank gil135225 sp P47730 DHAS_MYCBO - ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (ASA DEHYDROGENASE)		dehydrogenase	1009, 1010, 1034
321	52191619 (14103, 14104)	Novel Protein sim. GBank gil135560 sp P07913 TDH_ECOLI - THREONINE 3-DEHYDROGENASE		dehydrogenase	1040
322	27831633 (1837, 1838)	Novel Protein sim. GBank gil140564 sp P25377 YCZ5_YEAST - HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEIN IN PAU3 3 REGION		dehydrogenase	1022
323	34873651 (131, 132)	Novel Protein sim. GBank gil1480709 (U62057) - dihydrolipoprotein dehydrogenase [Mycoplasma capricolum]		dehydrogenase	1029
324	94134442 (133, 134)	Novel Protein sim. GBank gil15183540 (U53418) - UDP-glucose dehydrogenase [Glycine max]		dehydrogenase	1003

325	17902431 (21481, 21482)	Novel Protein sim. GBank gil1518540 (U53418) - UDP-glucose dehydrogenase [Glycine max]		dehydrogenase	1023
326	70336330 (19387, 19388)	Novel Protein sim. GBank gil1542914 emb CAB02185 - (Z80108) fmt [Mycobacterium tuberculosis]		dehydrogenase	1009
327	78204435 (6867, 6868)	Novel Protein sim. GBank gil1552860 emb CAB02456 - (Z80343) hypothetical protein RV3777 [Mycobacterium tuberculosis]		dehydrogenase	1026
328	25137971 (2797, 2798)	Novel Protein sim. GBank gil1552545 (U37486) - peroxisomal multifunctional enzyme type II [Rattus norvegicus]		dehydrogenase	1003
329	10106291 (2397, 2398)	Novel Protein sim. GBank gil1652995 dbj BAA17912 - (D90910) IMP dehydrogenase subunit [Synechocystis sp.]		dehydrogenase	1012
330	80502360 (17857, 17858)	Novel Protein sim. GBank gil1653603 dbj BAA18516 - (D90914) P protein of glycine cleavage complex [Synechocystis sp.]		dehydrogenase	1012
331	20284995 (14513, 14514)	Novel Protein sim. GBank gil1703381 sp P54894 ARGC_ANASP - N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GLUTAMATE SEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE)		dehydrogenase	1034
332	86691862 (13623, 13624)	Novel Protein sim. GBank gil1703381 sp P54895 ARGC_STRCO - N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (AGPR) (N-ACETYL-GLUTAMATE SEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE)		dehydrogenase	1025, 1053
333	79468408 (6801, 6802)	Novel Protein sim. GBank gil1706416 sp P42974 DHNA_BACSU - NADH DEHYDROGENASE (ALKYL HYDROPEROXIDE REDUCTASE)		dehydrogenase	1022

334	20177093 (20575, 20576)	Novel Protein sim. GBank gi 1708441 sp P54533 DLD2_BACSU - LIPOAMIDE DEHYDROGENASE COMPONENT OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (E3) (DIHYDROLIPOAMIDE DEHYDROGENASE) (LPD-V(Al))		dehydrogenase	1010
335	20613284 (20905, 20906)	Novel Protein sim. GBank gi 1708389 sp P50216 IDH_CORG1 - ISOCITRATE DEHYDROGENASE (NADP) (OXALOSUCCINATE DECARBOXYLASE) (IDH)		dehydrogenase	1022
336	86662499 (917, 918)	Novel Protein sim. GBank gi 1708408 sp P16100 IDH_AZ0V1 - ISOCITRATE DEHYDROGENASE (NADP) (OXALOSUCCINATE DECARBOXYLASE) (IDH)		dehydrogenase	1019
337	79483276 (11097, 11098)	Novel Protein sim. GBank gi 1708478 sp P500941 MH3_YEAST - PROBABLE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMP DHYDROGENASE) (IMPDH) (IMPD)		dehydrogenase	1006
338	78785362 (12953, 12954)	Novel Protein sim. GBank gi 1709183 sp Q11148 MURB_MYCTU - PUTATIVE UDP-N-ACETYLGLUCOSAMINE REDUCTASE (UDP-N-ACETYLGLUCOSAMINE DEHYDROGENASE)		dehydrogenase	1017
339	85800701 (10011, 10012)	Novel Protein sim. GBank gi 1709445 sp Q10504 ODP1_MYCTU - PYRUVATE DEHYDROGENASE E1 COMPONENT		dehydrogenase	1010, 1023, 1024, 1025
340	80219672 (13497, 13498)	Novel Protein sim. GBank gi 171785 M26390 - alpha-ketoglutarate dehydrogenase [Saccharomyces cerevisiae]		dehydrogenase	1003, 1006, 1008, 1010, 1017, 1022, 1024, 1026, 1034

341	80258034 (4465, 4466)	Novel Protein sim. GBank gil 17237631 sim P53315 YGS5A_ YEAST - MGA1 INTERGENIC REGION		dehydrogenase	1003, 1017, 1022, 1023, 1025, 1044
342	80226910 (7209, 7210)	Novel Protein sim. GBank gil 1736707 dbj BAA15860 - (D90840) UDG protein, [Escherichia coli]		dehydrogenase	1008, 1024
343	10129786 (14427, 14428)	Novel Protein sim. GBank gil 1763227 (U72152) - erythrose-4-phosphate dehydrogenase [Vibrio cholerae]		dehydrogenase	1044
344	85522996 (16739, 16740)	Novel Protein sim. GBank gil 1781301 emb CAA65229 - (X95385) glucose dehydrogenase (pyrroloquinoline-quinone) [Pantoea citrea]		dehydrogenase	1010, 1049
345	39711823 (331, 332)	Novel Protein sim. GBank gil 1805393 dbj BAA08955 - (D50453) 68% identity protein to 1-pyrroline-5-carboxylate dehydrogenase of B. subtilis [Bacillus subtilis]		dehydrogenase	1038
346	11093680 (21543, 21544)	Novel Protein sim. GBank gil 1805460 dbj BAA09022 - (D50453) homologue of succinate semialdehyde dehydrogenase GabD of E. coli [Bacillus subtilis]		dehydrogenase	1022
347	30873278 (1157, 1158)	Novel Protein sim. GBank gil 1806234 emb CAB06498 - (Z84498) hypothetical protein Rv1928c [Mycobacterium tuberculosis]		dehydrogenase	1008
348	29893414 (3499, 3500)	Novel Protein sim. GBank gil 1806234 emb CAB06498 - (Z84498) hypothetical protein Rv1928c [Mycobacterium tuberculosis]		dehydrogenase	1008
349	20287836 (5151, 5152)	Novel Protein sim. GBank gil 1877028 dbj BAA12222 - (D84102) 2-oxoglutarate dehydrogenase [Corynebacterium glutamicum]		dehydrogenase	1034

350	86471959 (7923, 7924)	Novel Protein sim. GBank gj 1877028[dbj BAA12222] - (D84102) 2'-oxoglutarate dehydrogenase [Corynebacterium glutamicum]		dehydrogenase	1029, 1048
351	79574453 (4109, 4110)	Novel Protein sim. GBank gj 1890604[embj CAB01952] - (Z79692) ExpA10 [Sinorhizobium meliloti]		dehydrogenase	1031
352	17695549 (10435, 10436)	Novel Protein sim. GBank gj 1938240[embj CAA71232] - (Y10142) complex I 51kDa subunit [Rhodobacter capsulatus]		dehydrogenase	1044
353	20606791 (2129, 2130)	Novel Protein sim. GBank gj 2113942[embj CAB08927] - (Z95556) hypothetical protein Rv2509 [Mycobacterium tuberculosis]		dehydrogenase	1004
354	20467370 (6113, 6114)	Novel Protein sim. GBank gj 2127114[pirj S66039 - IMP dehydrogenase - Bacillus subtilis]		dehydrogenase	1010
355	11340655 (12543, 12544)	Novel Protein sim. GBank gj 2131327[pirj S67621 - hypothetical protein YDL085w - yeast (Saccharomyces cerevisiae)]		dehydrogenase	1004, 1010
356	78778423 (9815, 9816)	Novel Protein sim. GBank gj 2144331[pirj S60929 - probable aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) YPL061w - yeast (Saccharomyces cerevisiae)]		dehydrogenase	1009, 1026
357	79561728 (22653, 22654)	Novel Protein sim. GBank gj 2228233 (U24215) - p-curmic aldehyde dehydrogenase [Pseudomonas putida]		dehydrogenase	1044
358	27849244 (19153, 19154)	Novel Protein sim. GBank gj 2228735 (U89767) - glutathione dependent formaldehyde dehydrogenase [Anabaena azollae]		dehydrogenase	1022
359	20369215 (2907, 2908)	Novel Protein sim. GBank gj 2313134[gb AAD07126.11 - (AE000527) delta-1-pyrroline-5-carboxylate dehydrogenase [Helicobacter pylori 26695]		dehydrogenase	1004

360	78952543 (919, 920)	Novel Protein sim. GBank g 231955 sp P30234 DHA_MYCTU - ALANINE DEHYDROGENASE (40 KD ANTIGEN)		dehydrogenase	1016
361	81758941 (6327, 6328)	Novel Protein sim. GBank g 2326687 emb CA810998 - (Z98271) methylenetetrahydrotetolide dehydrogenase [Mycobacterium leprae]		dehydrogenase	1004, 1006, 1009, 1010, 1012, 1024, 1030, 1031, 1034, 1054
362	9893326 (2777, 2778)	Novel Protein sim. GBank g 2360965 (AF016253) - D-amino acid dehydrogenase [Klebsiella aerogenes]		dehydrogenase	1034
363	79406542 (11679, 11680)	Novel Protein sim. GBank g 2443238 (dbj BAA22309 - (D86417) YfIP [Bacillus subtilis])		dehydrogenase	1016
364	11269485 (9989, 9990)	Novel Protein sim. GBank g 2492777 sp Q04884 YM97 - YEAST - HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEIN IN PRE5-FET4 INTERGENIC REGION		dehydrogenase	1006
365	79491842 (19933, 19934)	Novel Protein sim. GBank g 2494074 sp P55653 GABD_RHISN - PROBABLE SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (SSDH)		dehydrogenase	1024
366	9897086 (4149, 4150)	Novel Protein sim. GBank g 2495628 sp P55757 YOHL_SERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BOA 5'REGION		dehydrogenase	1039
367	80430354 (2899, 2900)	Novel Protein sim. GBank g 2497145 sp Q04304 YMYO - YEAST - HYPOTHETICAL 24.9 KD PROTEIN IN RCA1-NPL6 INTERGENIC REGION		dehydrogenase	1001, 1003, 1006, 1008, 1009, 1010, 1012, 1017, 1022, 1023, 1024, 1026, 1027, 1030, 1034, 1038, 1039, 1044

368	11088223 (4931, 4932)	Novel Protein sim. GBank gi 2497360 sp Q50715 IMDH_MYCTU - INOSINE-5-MONOPHOSPHATE DEHYDROGENASE (IMPDH) (IMPDH)		dehydrogenase	1004
369	20724157 (14613, 14614)	Novel Protein sim. GBank gi 249312 sp Q92406 NUBM_ASPPNG - NADH-UBIQUINONE OXIDOREDUCTASE 51. KD SUBUNIT PRECURSOR (COMPLEX I-51KD) (CI-51KD)		dehydrogenase	1006
370	30670264 (3981, 3982)	Novel Protein sim. GBank gi 2498411 sp Q59637 ODP1_PSEAE - PYRUVATE DEHYDROGENASE E1 COMPONENT		dehydrogenase	1026
371	78466424 (809, 810)	Novel Protein sim. GBank gi 2498419 sp Q508601 GCSP_MYCTU - PROBABLE GLYCINE DEHYDROGENASE (DECARBOXYLATING) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)		dehydrogenase	1026
372	70945116 (13889, 13890)	Novel Protein sim. GBank gi 2498419 sp Q508601 GCSP_MYCTU - PROBABLE GLYCINE DEHYDROGENASE (DECARBOXYLATING) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)		dehydrogenase	1029
373	20727915 (7429, 7430)	Novel Protein sim. GBank gi 2578379 emb CAA15464 - (AL008609) glycine dehydrogenase (decarboxylating) [Mycobacterium leprae]		dehydrogenase	1006
374	80062798 (7533, 7534)	Novel Protein sim. GBank gi 2645697 (AF031940) - alcohol dehydrogenase [Sinorhizobium meliloti]		dehydrogenase	1024
375	95005632 (13757, 13758)	Novel Protein sim. GBank gi 2645697 (AF031940) - alcohol dehydrogenase [Sinorhizobium meliloti]		dehydrogenase	1001, 1004, 1012, 1023, 1025, 1029, 1034

376	66396374 (15967, 15968)	Novel Protein sim. GBank gi 2645697 (AF031940) - alcohol dehydrogenase [Sinorhizobium meliloti]		dehydrogenase	1008
377	20405820 (1961, 1962)	Novel Protein sim. GBank gi 2695834 emb CAA15904 - (AL021006) suCA [Mycobacterium tuberculosis]		dehydrogenase	1004, 1034
378	11708857 (7977, 7978)	Novel Protein sim. GBank gi 2695834 emb CAA15904 - (AL021006) suCA [Mycobacterium tuberculosis]		dehydrogenase	1016
379	66317781 (8695, 8696)	Novel Protein sim. GBank gi 2695834 emb CAA15904 - (AL021006) suCA [Mycobacterium tuberculosis]		dehydrogenase	1009
380	56716390 (8025, 8026)	Novel Protein sim. GBank gi 2782310 (AF040570) - unknown [Amycolatopsis mediterranei]		dehydrogenase	1022
381	80500722 (9773, 9774)	Novel Protein sim. GBank gi 281569 pir S27612 - ketoglutarate semialdehyde dehydrogenase - Pseudomonas putida		dehydrogenase	1010, 1012
382	79833299 (11977, 11978)	Novel Protein sim. GBank gi 281570 pir S27619 - ketoglutarate semialdehyde dehydrogenase - Pseudomonas putida		dehydrogenase	1023
383	79858B33 (14713, 14714)	Novel Protein sim. GBank gi 2829734 sp P94132 ETFD_ACICA - PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-UBIQUINONE OXIDOREDUCTASE (ETF-QO) (ETF-UBIQUINONE OXIDOREDUCTASE) (ETF-DEHYDROGENASE) (ELECTRON-TRANSFERRING-FLAVOPROTEIN DEHYDROGENASE)		dehydrogenase	1039
384	55286916 (21383, 21384)	Novel Protein sim. GBank gi 2851634 sp Q50591 YOD1_MYCTU - HYPOTHETICAL 50.0 KD PROTEIN CY1A11.01		dehydrogenase	1029, 1053

385	79490301 (17189, 17190)	Novel Protein sim. GBank [gi 287897 emb CAA22649_1] - (X60046) orfx [Oenothera berteriana]		dehydrogenase	1008, 1022
386	20443141 (10455, 10456)	Novel Protein sim. GBank [gi 2894252 emb CAA17114_1] - (AL021841) hypothetical protein RV3342 [Mycobacterium tuberculosis]		dehydrogenase	1024
387	16331388 (14923, 14924)	Novel Protein sim. GBank [gi 2895866 (AF045770) - methylmalonate semi-aldehyde dehydrogenase [Oryza sativa]		dehydrogenase	1006
388	94140370 (7589, 7590)	Novel Protein sim. GBank [gi 2911027 emb CAA17520] - (AL021958) mmssA [Mycobacterium tuberculosis]		dehydrogenase	1010, 1049
389	94146368 (11449, 11450)	Novel Protein sim. GBank [gi 2911027 emb CAA17520] - (AL021958) mmssA [Mycobacterium tuberculosis]		dehydrogenase	1003
390	20293077 (14295, 14296)	Novel Protein sim. GBank [gi 2911027 emb CAA17520] - (AL021958) mmssA [Mycobacterium tuberculosis]		dehydrogenase	1034
391	18561962 (15221, 15222)	Novel Protein sim. GBank [gi 2911042 emb CAA17552] - (AL021961) Phosphoglycerate dehydrogenase - like protein [Arabidopsis thaliana]		dehydrogenase	1017, 1025
392	27981868 (21995, 21996)	Novel Protein sim. GBank [gi 2947296 (AF049107) - alanine dehydrogenase [Myxococcus xanthus]		dehydrogenase	1006
393	78736706 (16755, 16756)	Novel Protein sim. GBank [gi 2982328 (AF051249) - pyruvate dehydrogenase E1 beta subunit [Picea mariana]		dehydrogenase	1017
394	20623579 (21611, 21612)	Novel Protein sim. GBank [gi 3025211 sp P75804 YLI_ECOLI - HYPOTHETICAL 41.1 KD PROTEIN IN MOEA-DACC INTERGENIC REGION PRECURSOR		dehydrogenase	1004
395	79485865 (9205, 9206)	Novel Protein sim. GBank [gi 3025295 sp Q46856 YQHD_ECOLI - HYPOTHETICAL OXIDOREDUCTASE IN METC-SUFI INTERGENIC REGION		dehydrogenase	1022

396	86094813 (11485, 11486)	Novel Protein sim. GBank gj 3088561 (AF059313) - myo-inositol dehydrogenase [Sinorhizobium meliloti]		dehydrogenase	1004, 1017, 1022, 1031
397	27837984 (7415, 7416)	Novel Protein sim. GBank gj 312179 emb CAA51676 - (X73151) glyceraldehyde 3-phosphate dehydrogenase (phosphorylating) [Zea mays]		dehydrogenase	1008
398	79641344 (5357, 5358)	Novel Protein sim. GBank gj 313597 emb CAA17432 - (AL021939) dj352A20.2 (aldehyde dehydrogenase family protein) [Homo sapiens]		dehydrogenase	1003, 1009, 1025
399	28375955 (9075, 9076)	Novel Protein sim. GBank gj 3150233 emb CAA19205 - (AL023635) pyruvate dehydrogenase e1 component [Mycobacterium leprae]		dehydrogenase	1016
400	80105552 (21271, 21272)	Novel Protein sim. GBank gj 3219721 (AF068740) - alpha-ketoglutarate dehydrogenase; E1 [Pseudomonas putida]		dehydrogenase	1013
401	28386820 (18003, 18004)	Novel Protein sim. GBank gj 324228 emb CAA16668 - (AL021646) nuoM [Mycobacterium tuberculosis]		dehydrogenase	1044
402	39567937 (14075, 14076)	Novel Protein sim. GBank gj 333420 isp Q49954 GCSP - SOLTU - GLYCINE DEHYDROGENASE (DECARBOXYLATING) PRECURSOR (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN)		dehydrogenase	1006
403	25264319 (4211, 4212)	Novel Protein sim. GBank gj 3402239 emb CAA20175 - (AL031184) aminomethyltransferase [Streptomyces coelicolor]		dehydrogenase	1008
404	94672537 (13741, 13742)	Novel Protein sim. GBank gj 3746332 (AF016307) - possible NADH-dependent oxidase, may function as a demethylase [Sinorhizobium meliloti]		dehydrogenase	1022

405	29273027 (5121, 5122)	Novel Protein sim. GBank gi 3851005 gb AAC72195 - (AF069911) pyruvate dehydrogenase E1 alpha subunit [Zea mays]		dehydrogenase	1022
406	11703346 (20947, 20948)	Novel Protein sim. GBank gi 38683 emb CAA37101 - (X52913) nodG protein (AA 1-254) [Azospirillum brasiliense]		dehydrogenase	1031
407	11088473 (17339, 17340)	Novel Protein sim. GBank gi 3888720 (AE000464) - formate dehydrogenase-O, major subunit [Escherichia coli]		dehydrogenase	1004
408	57458947 (20489, 20490)	Novel Protein sim. GBank gi 3879884 emb CAA98524 - (Z74042) predicted using Genefinder; Similarity to Haemophilus 3-oxoacyl-(acyl-carrier protein) reductase (SW: FABG_HAEIN); cDNA EST yk470b2.3 comes from this gene; cDNA EST yk470b2.5 comes from this gene [Caenorhabditis elegans]		dehydrogenase	1011
409	54700472 (5107, 5108)	Novel Protein sim. GBank gi 3881461 emb CAA92990.1 - (Z68753) predicted using Genefinder; Similarity to Glucose-repressible alcohol dehydrogenase transcriptional effector (SW:CCR4_YEAST); cDNA EST EMBL:T01851 comes from this gene; cDNA EST yk240e8.5 comes from this gene; cDNA ...		dehydrogenase	1041
410	30256630 (567, 568)	Novel Protein sim. GBank gi 3913464 sp Q24174 DHAB_ORYSA - BETAINE-ALDEHYDE DEHYDROGENASE (BADH)		dehydrogenase	1024
411	20177067 (16745, 16746)	Novel Protein sim. GBank gi 3914051 sp O69282 MQO_CORGL - MALATE:QUINONE OXIDOREDUCTASE (MALATE DEHYDROGENASE (ACCEPTOR)) (MQO)		dehydrogenase	1010

412	12646950 (13495, 13496)	Novel Protein sim. GBank gi 3990091 spl P31975 ADH_MYCTU - NADP- DEPENDENT ALCOHOL DEHYDROGENASE		dehydrogenase	1006
413	10047517 (20801, 20802)	Novel Protein sim. GBank gi 39939 spl P31046 LDL3_PSEPU - DIHYDROLIPOAMIDE DEHYDROGENASE 3 (LD-3)		dehydrogenase	1039
414	21431110 (1631, 1632)	Novel Protein sim. GBank gi 4007718 emb CAA22402 - (AL034447) glycerol- 3-phosphate dehydrogenase [Streptomyces coelicolor]		dehydrogenase	1004
415	17854237 (14821, 14822)	Novel Protein sim. GBank gi 4098682 (U80806) - CutL [Hydrogenophaga_pseudoflava]		dehydrogenase	1003
416	79578900 (16037, 16038)	Novel Protein sim. GBank gi 417432 spl Q03558 OYE2 - YEAST - NADPH DEHYDROGENASE 2 (OLD YELLOW ENZYME 2)		dehydrogenase	1010
417	20440650 (9447, 9448)	Novel Protein sim. GBank gi 4204381 (U73127) - 2-hydroxymuconic semialdehyde dehydrogenase [Sphingomonas sp. A8AN3]		dehydrogenase	1024
418	27370766 (4271, 4272)	Novel Protein sim. GBank gi 4210630 emb CAA76340 - (Y16642) dihydrolipoamide dehydrogenase [Corynebacterium glutamicum]		dehydrogenase	1003
419	79852628 (5795, 5796)	Novel Protein sim. GBank gi 4240419 (AF060235) - oxidoreductase homolog [Streptomyces cyanogenus]		dehydrogenase	1003, 1017
420	55281868 (7075, 7076)	Novel Protein sim. GBank gi 4321580 gb AAD15785 - (AF050114) alginat- lyase [Pseudomonas sp. W7]		dehydrogenase	1029, 1053
421	20459744 (14395, 14396)	Novel Protein sim. GBank gi 4321580 gb AAD15785 - (AF050114) alginat- lyase [Pseudomonas sp. W7]		dehydrogenase	1010
422	11687220 (9873, 9874)	Novel Protein sim. GBank gi 436913 (U01158) - nicotinamide nucleotide transhydrogenase, subunit alpha 1 [Rhodospirillum rubrum]		dehydrogenase	1031

423	80028168 (22831, 22832)	Novel Protein sim. GBank gi 4377316 gb AAD19125 - (AE001679) UDP-N- Acetylglucosaminide acetyltransferase [Chlamydia pneumoniae]		dehydrogenase	1006, 1022, 1027
424	78901322 (8159, 8160)	Novel Protein sim. GBank gi 4505493 ref NP_002532.1 pOGDH - oxoglutarate dehydrogenase (lipoyamide) electrotransfer ubiquinone oxidoreductase		dehydrogenase	1003
425	82348699 (7109, 7110)	Novel Protein sim. GBank gi 4511983 gb AAD21543.1 - (AF088896) [Zymomonas mobilis]		dehydrogenase	1001, 1010, 1012, 1022
426	78727367 (12275, 12276)	Novel Protein sim. GBank gi 4557343 ref NP_001173.1 pATTQ1 - antiquitin 1		dehydrogenase	1008
427	79447205 (9267, 9268)	Novel Protein sim. GBank gi 4577412 gb AAD23901.1 AF00941 - (AF009415) choline dehydrogenase [Staphylococcus xylosus]		dehydrogenase	1006
428	79834660 (6641, 6642)	Novel Protein sim. GBank gi 4585638 emb CAB40932.1 - (AL049630) putative NADH dehydrogenase [Streptomyces coelicolor]		dehydrogenase	1010, 1016, 1023
429	29013008 (3667, 3668)	Novel Protein sim. GBank gi 4621971 sp Q00055 GPD1_YEAST - GLYCEROL-3-PHOSPHATE DEHYDROGENASE (NAD+ 1)		dehydrogenase	1039
430	112911700 (9561, 9562)	Novel Protein sim. GBank gi 481230 pir S38363 - L-iditol 2-dehydrogenase (EC 1.1.1.14) precursor - rat		dehydrogenase	1010
431	16716670 (10165, 10166)	Novel Protein sim. GBank gi 4829856 ref NP_004997.1 pNDUF - NADH dehydrogenase (ubiquinone) Fe-S protein 1 (75KD) (NADH-coenzyme Q reductase)		dehydrogenase	1017
432	20715202 (15599, 15600)	Novel Protein sim. GBank gi 48848855 gb AAD31841.1 AF13275 - (AF132754) coenzyme A acylating aldehyde dehydrogenase [Clostridium beijerinckii]		dehydrogenase	1022

433	100326400 (22123, 222124)	Novel Protein sim. G-Bank gi 4885387 refNP_005318.1 pHADH - L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain		dehydrogenase	1058
434	10104354 (11435, 11436)	Novel Protein sim. G-Bank gi 4887168 gb AAD32214.1 - (AF140745) sarcosine dehydrogenase [Homo sapiens]		dehydrogenase	1003
435	70287085 (6721, 6722)	Novel Protein sim. G-Bank gi 4980917 gb AAD35497.1 AE00172 - (AE001720) alcohol dehydrogenase, zinc-containing [Thermotoga maritima]		dehydrogenase	1009
436	80068254 (3241, 3242)	Novel Protein sim. G-Bank gi 4981559 gb AAD36088.1 AE00176 - (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]		dehydrogenase	1010
437	10106020 (18189, 18190)	Novel Protein sim. G-Bank gi 5042223 emb CAB44527.1 - (AL078618) nuoE, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	1012
438	39563115 (14565, 14566)	Novel Protein sim. G-Bank gi 531467 (U12977) - similar to glycerol-3-phosphate-dehydrogenase, GenBank Accession Number U00039, and Swiss-Prot Accession Number P-13706 [Pseudomonas larmigniei]		dehydrogenase	1006
439	19887534 (20437, 20438)	Novel Protein sim. G-Bank gi 5508821 gb AAD43988.1 U59485_- (U59485)		dehydrogenase	1034
440	85800229 (4063, 4064)	Novel Protein sim. G-Bank gi 5578883 emb CAE51265.1 - (AL096872) putative dihydrolipoamide succinyltransferase [Streptomyces coelicolor A3(2)]		dehydrogenase	1025
441	80080560 (21335, 21336)	Novel Protein sim. G-Bank gi 5578865 emb CAE51267.1 - (AL096872) putative pyruvate dehydrogenase E1 component [Streptomyces coelicolor A3(2)]		dehydrogenase	1006, 1034

442	78731548 (22533, 22534)	Novel Protein sim. GBank gil566903 gb AAD46162_1 AF13030 - (AF130307) glutathione-dependent formaldehyde dehydrogenase [Acinetobacter baumannii]		dehydrogenase	1008
443	78468748 (9061, 9062)	Novel Protein sim. GBank gil568925 dbj BAA82881_1 - (AB024335) similar to orf5 [Comamonas testosteroni]		dehydrogenase	1026
444	20628080 (14887, 14888)	Novel Protein sim. GBank gil568920 dbj BAA82881_1 - (AB024335) similar to orf5 [Comamonas testosteroni]		dehydrogenase	1010
445	10857466 (6007, 6008)	Novel Protein sim. GBank gil568906 emb CAB52003_1 - (AL109663) putative regulatory protein [Streptomyces coelicolor A3(2)]		dehydrogenase	1006
446	19740397 (19749, 19750)	Novel Protein sim. GBank gil584729 sp P3811_3 ADH5_YEAST - ALCOHOL DEHYDROGENASE_V		dehydrogenase	1001
447	27978011 (8763, 8764)	Novel Protein sim. GBank gil585047 sp Q08352 DHA_BACSU - ALANINE DEHYDROGENASE (STAGE V SPORULATION PROTEIN N)		dehydrogenase	1006
448	32153300 (15259, 15260)	Novel Protein sim. GBank gil585952 sp P3810_5 RSPB_ECOLI - STARVATION SENSING PROTEIN RSPB		dehydrogenase	1039
449	65704010 (15355, 15356)	Novel Protein sim. GBank gil5863_3 sp P20049 TYR1_YEAST - PREPHENATE DEHYDROGENASE (NADP+) (PRDH)		dehydrogenase	1010
450	19883381 (6817, 6818)	Novel Protein sim. GBank gil586147 sp P37791 UDG_SHIFL - UDP-GLUCOSE 6-DEHYDROGENASE (UDP-GLC DEHYDROGENASE)		dehydrogenase	1001
451	79184302 (14519, 14520)	Novel Protein sim. GBank gil628685 pir S40838 - fdG protein - Escherichia coli		dehydrogenase	1027

452	54528091 (21925, 21926)	Novel Protein sim. GBank gi 629429 pir S43340 - glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) (EC 1.2.1.13) - red alga (Chondrus crispus)		dehydrogenase	1041
453	56998990 (4067, 4068)	Novel Protein sim. GBank gi 655978 pir DDEECNC - Glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) (anaerobic) chain C - Escherichia coli		dehydrogenase	1026
454	78463737 (2595, 2596)	Novel Protein sim. GBank gi 699253 (J15183) - lipoamide dehydrogenase [Mycobacterium leprae]		dehydrogenase	1026
455	20289183 (7809, 7810)	Novel Protein sim. GBank gi 728805 sp P39450 ADH3_PASPI - PUTATIVE ALCOHOL DEHYDROGENASE CLASS III (GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE) (FDH) (FALDH)		dehydrogenase	1034
456	80080812 (7647, 7648)	Novel Protein sim. GBank gi 7298614 sp P3218 GPD_M_YEAST - GLYCEROL-3-PHOSPHATE DEHYDROGENASE MITOCHONDRIAL PRECURSOR (GPD-M) (GPDH-M)		dehydrogenase	1034
457	78768789 (8249, 8250)	Novel Protein sim. GBank gi 731731 sp P38858 SOL3_YEAST - SOL3 PROTEIN		dehydrogenase	1026
458	52563171 (1559, 1560)	Novel Protein sim. GBank gi 732112 sp P39400 YJJN_ECOLI - HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEIN IN TSR-MDOB INTERGENIC REGION		dehydrogenase	1024
459	24129630 (14077, 14078)	Novel Protein sim. GBank gi 732112 sp P39400 YJJN_ECOLI - HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEIN IN TSR-MDOB INTERGENIC REGION		dehydrogenase	1003, 1022

460	20385694 (22043, 22044)	Novel Protein sim. GBank gi 732112 sp P39400 YJIN_ECOLI - HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEIN IN TSR-MDOB INTERGENIC REGION		dehydrogenase	1022
461	559977098 (7299, 7300)	Novel Protein sim. GBank gi 746454 (U23510) - similar to glycine dehydrogenase [Caenorhabditis elegans]		dehydrogenase	1026
462	65488254 (15403, 15404)	Novel Protein sim. GBank gi 757827 emb CAA860411 - (Z37980) 5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase [Escherichia coli]		dehydrogenase	1023
463	79186444 (20983, 20984)	Novel Protein sim. GBank gi 78302 pif JU0397 - probable aldehyde dehydrogenase (EC 1.2.-.-) - Escherichia coli		dehydrogenase	1027
464	78377278 (17035, 17036)	Novel Protein sim. GBank gi 913631 bbs 163051 - class-3 aldehyde dehydrogenase, class-3 ALDH [Synechococcus, PC7942, Peptide, 459 aa]		dehydrogenase	1026
465	39523788 (15189, 15190)			dehydrogenase	1004
466	21428995 (17251, 17252)			dehydrogenase	1006
467	86473407 (11889, 11890)	Novel Protein sim. GBank gi 134944 sp P22243 STAD_CART1 - ACYL-[ACYL-CARRIER PROTEIN] DESATURASE PRECURSOR (STEAROYLACP DESATURASE)	Contains protein domain (PF00487) - Fatty acid desaturase	desaturase	1006, 1024, 1044, 1054
468	28811447 (14733, 14734)	Novel Protein sim. GBank gi 3341398 emb CAA76158 - (Y16285) delta 12 fatty acid acyl/enase [Crepis alpina]	Contains protein domain (PF00487) - Fatty acid desaturase	desaturase	1009
469	79610725 (7927, 7928)	Novel Protein sim. GBank gi 1055272 (U37517) - stearoyl-CoA desaturase [Azotobacter vinelandii]		desaturase	1004
470	10814277 (22759, 22760)	Novel Protein sim. GBank gi 2197201 (U36390) - delta-9 desaturase [Synechococcus PCC7002]		desaturase	1017

471	80236986 (3685 3686)	Novel Protein sim. GBank gi 416963 sp P32353 ERG3_YEAST - C-5 STEROL DESATURASE		desaturase	1003, 1004, 1006, 1008, 1017, 1022, 1023, 1039, 1044
472	9892172 (13831, 13832)	Novel Protein sim. GBank gi 113629 sp P23747 ALGB_PSEAE - ALGINATE BIOSYNTHESIS TRANSCRIPTIONAL REGULATORY PROTEIN ALGB	Contains protein domain (PF00072) - Response regulator receiver domain	dna_ma_bind	1034
473	80503688 (4643, 4644)	Novel Protein sim. GBank gi 1723676 sp P53243 YG2A_YEAST - PUTATIVE 91.0 KD ZINC FINGER PROTEIN IN SPT4-ROM1 INTERGENIC REGION	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1003, 1006, 1012, 1022
474	94630880 (6115, 6116)	Novel Protein sim. GBank gi 19418 sp P0db 1MEY C - Chain C, Crystal Structure Of A Designed Zinc Finger Protein Bound To Dna	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1008, 1040, 1054
475	80177716 (15997, 15998)	Novel Protein sim. GBank gi 3477297 (AC002310) - Unknown gene product [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1022
476	28987727 (8757, 8758)	Novel Protein sim. GBank gi 543209 pir S41688 - DNA-binding protein - mouse	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1008
477	79922130 (2009, 2010)	Novel Protein sim. GBank gi 543346 pir S42077 - zinc finger protein 30 - mouse	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1003
478	27977944 (13183, 13184)	Novel Protein sim. GBank gi 545417 6ref NP_008376.1 pZNF - zinc finger protein C2H2-25	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_ma_bind	1006
479	79444824 (16141, 16142)	Novel Protein sim. GBank gi 1138833 sp P12626 ANFA_AZQVI - NITROGEN FIXATION PROTEIN ANFA	Contains protein domain (PF00158) - Sigma-54 transcription factors	dna_ma_bind	1034
480	20481416 (21357, 21358)	Novel Protein sim. GBank gi 3228542 (AF061028) - HrpR [Pseudomonas syringae pv. tomato]	Contains protein domain (PF00158) - Sigma-54 transcription factors	dna_ma_bind	1022

481	11103232 (22723, 22724)	Novel Protein sim. GBank gi 19168865 (U85412) - response regulator BpdT [Rhodococcus sp. M5]	Contains protein domain (PF00196) - Bacterial regulatory proteins, luxR family	dna_rna_bind	1022
482	79420904 (4551, 4552)	Novel Protein sim. GBank gi 731913 sp P40569 YIT3_YEAST - HYPOTHETICAL 132 KD PROTEIN IN DBF8-MET28 INTERGENIC REGION	Contains protein domain (PF00320) - GATA zinc finger	dna_rna_bind	1006
483	80087172 (4739, 4740)	Novel Protein sim. GBank gi 586039 sp P37455 SSB_BACSU - SINGLE-STRAND BINDING PROTEIN (SSB) (HELIX-DESTABILIZING PROTEIN)	Contains protein domain (PF00436) - Single-strand binding protein family	dna_rna_bind	1034
484	79242051 (7225, 7226)	Novel Protein sim. GBank gi 5420016 emb CA846398_1 - (AL096743) putative transcriptional regulator [Streptomyces coelicolor]	Contains protein domain (PF00456) - Bacterial regulatory proteins, deoR family	dna_rna_bind	1003, 1044
485	80080168 (19499, 19500)	Novel Protein sim. GBank gi 584887 sp P38042 CC27_YEAST - CELL DIVISION CONTROL PROTEIN 27	Contains protein domain (PF00515) - TPR Domain	dna_rna_bind	1006, 1009, 1017, 1034
486	65496651 (17395, 17396)	Novel Protein sim. GBank gi 2564960 (L13845) - DNA-binding protein; ORF3; putative [Sinorhizobium meliloti]	Contains protein domain (PF01381) - Helix-turn-helix	dna_rna_bind	1023
487	10030229 (3447, 3448)	Novel Protein sim. GBank gi 106323 pir A34087 - hypothetical protein (L1H 5' region) - human		dna_rna_bind	1038
488	33072745 (8957, 8958)	Novel Protein sim. GBank gi 106323 pir A34087 - hypothetical protein (L1H 5' region) - human		dna_rna_bind	1016
489	8487084 (15907, 15908)	Novel Protein sim. GBank gi 114834 sp P22035 BAS1_YEAST - MYB-LIKE DNA-BINDING PROTEIN BAS1		dna_rna_bind	1022

490	80048279 (17691, 17692)	Novel Protein sim. GBank gjl131639 sp P15639 PUR9_ECOLI - PHOSPHORIBOSYLAMINOIMIDAZOLECARBOX AMIDE FORMYLTRANSFERASE (AICAR TRANSFORMYLASE) / IMP CYCLOHYDROLASE (INOSINICASE) (IMP SYNTHETASE) (ATIC)		dna_rna_bind	1006, 1010, 1016, 1022, 1024, 1034
491	80221007 (1391, 1392)	Novel Protein sim. GBank gjl132475 sp P26754 RFA2_YEAST - REPLICATION FACTOR-A PROTEIN 2 (RF-A) (DNA BINDING PROTEIN BUFI) (REPLICATION PROTEIN A 36 KD SUBUNIT)		dna_rna_bind	1009, 1010, 1022
492	78747826 (8799, 8800)	Novel Protein sim. GBank gjl134851 sp P06843 SPT2_YEAST - SPT2 PROTEIN (NEGATIVE REGULATOR OF TY TRANSCRIPTION)		dna_rna_bind	1022
493	85801306 (22631, 22632)	Novel Protein sim. GBank gjl165395 dbj BAA18861 - (D90917) phosphoribosyl aminimidazole carboxy formyl formyltransferase [Synnechocystis sp.]		dna_rna_bind	1025
494	796211761 (22537, 22538)	Novel Protein sim. GBank gjl1709915 sp P38009 PU92_YEAST - PHOSPHORIBOSYLAMINOIMIDAZOLECARBOX AMIDE FORMYLTRANSFERASE 2 (AICAR TRANSFORMYLASE) / IMP CYCLOHYDROLASE (INOSINICASE) (IMP SYNTHETASE) (ATIC)		dna_rna_bind	1004
495	10328824 (12287, 12288)	Novel Protein sim. GBank gjl72438 (L01797) - RNA-binding protein [Saccharomyces cerevisiae]		dna_rna_bind	1025
496	65693678 (7983, 7984)	Novel Protein sim. GBank gjl1899188 (U90212) - DNA binding protein ACBF [Nicotiana tabacum]		dna_rna_bind	1010, 1049

497	25323726 (14015) 14016)	Novel Protein sim. GBank gi 3024025 sp Q04067 IF35_YEAST - EUKARYOTIC TRANSLATION INITIATION FACTOR 3 RNA-BINDING SUBUNIT (EIF-3 RNA- BINDING SUBUNIT) (EIF3 P33) (TRANSLATION INITIATION FACTOR EIF3, P33 SUBUNIT)		dna_ma_bind	1003
498	17706888 (15317), 15318)	Novel Protein sim. GBank gi 3367769 emb CAA20094 - (AL031155) putative two-component regulator [Streptomyces coelicolor].		dna_ma_bind	1017
499	77881967 (22825, 22826)	Novel Protein sim. GBank gi 382917 pf 1901303A - Leu zipper protein p40 [Homo sapiens].		dna_ma_bind	1021
500	52560369 (18341, 18342)	Novel Protein sim. GBank gi 418429 sp P3264 YE06_YEAST - HYPOTHETICAL 127.0 KD PROTEIN IN RAD24- BMH1 INTERGENIC REGION		dna_ma_bind	1004, 1006
501	80239693 (229, 230)	Novel Protein sim. GBank gi 23560 pin A47318 - RNA-binding protein Raly - mouse		dna_ma_bind	1004, 1010, 1017, 1023, 1024, 1029
502	29004930 (5385, 5386)	Novel Protein sim. GBank gi 4850285 emb CAB43041.1 - (AL049876) putative protein [Arabidopsis thaliana]		dna_ma_bind	1044
503	11818622 (6737, 6738)	Novel Protein sim. GBank gi 5052163 ref NP_005640.1 pTCF1 - transcription factor 17		dna_ma_bind	1031
504	79781387 (13959, 13960)	Novel Protein sim. GBank gi 730108 sp Q00539 NAM8_YEAST - NAM8 PROTEIN		dna_ma_bind	1003, 1008, 1009, 1017, 1022, 1024, 1026, 1039, 1044
505	20472968 (18357, 18358)	Novel Protein sim. GBank gi 729810 sp P41133 D3_MOUSE - DNA- BINDING PROTEIN INHIBITOR ID-3 (ID-LIKE PROTEIN INHIBITOR HLH 462)		dna_ma_bind_inhib	1010

506	65495436 (18379, 18380)	Novel Protein sim. GBank g 3015621 (AF035460) - low molecular weight heat shock protein precursor [Zea mays]	Contains protein domain (PF00011) - Hsp20/alpha crystallin family	eph	1023, 1036, 1039
507	94747283 (15131, 15132)	Novel Protein sim. GBank g 545100 bb5 142990 - Shb=Src homology 2 protein [mice, Peptide Partial, 309 aa]	Contains protein domain (PF00017) - Src homology domain 2	eph	1040
508	94747286 (15291, 15292)	Novel Protein sim. GBank g 545100 bb5 142990 - Shb=Src homology 2 protein [mice, Peptide Partial, 309 aa]	Contains protein domain (PF00017) - Src homology domain 2	eph	1010, 1021, 1030, 1054, 1058
509	19884720 (19949, 19950)	Novel Protein sim. GBank g 45392 p1emb CAB39858, 1 - (AL049497) putative long chain fatty acid coA ligase [Streptomyces coelicolor]	Contains protein domain (PF00039) - Fibronectin type I domain	eph	1034
510	4955990 (11437, 11438)	Novel Protein sim. GBank g 1504040 dbj BAA13219 - (D86983) similar to D.melanogaster peroxidase(U11052) [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	eph	1036
511	86381126 (11323, 11324)	Novel Protein sim. GBank g 1019910 (L40172) - JAK3 gene product [Mus musculus]	Contains protein domain (PF00069) - Euksaryotic protein kinase domain	eph	1053
512	20751138 (7663, 7664)	Novel Protein sim. GBank g 346967 pir A44876 - protein-tyrosine kinase (EC 2.7.1.112) kit, truncated receptor form - mouse	Contains protein domain (PF00069) - Euksaryotic protein kinase domain	eph	1022
513	29180944 (14331, 14332)	Novel Protein sim. GBank g 11388 1 sp P11859 ANGT_MOUSE - ANGIOTENSIN-NOGEN PRECURSOR	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	eph	1003
514	77799017 (10849, 10850)	Novel Protein sim. GBank g 1184675 (U29172) - AlgW [Pseudomonas aeruginosa]	Contains protein domain (PF00089) - Trypsin	eph	1051
515	80079381 (12479, 12480)	Novel Protein sim. GBank g 11623 sp P1942 ICH60_COXBU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN B)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1034, 1038

516	78764917 (12517, 12518)	Novel Protein sim. GBank gil1341021spl P08823 RUBA_WHEAT - RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT PRECURSOR (60 KD CHAPERONIN ALPHA SUBUNIT) (CPN-60 ALPHA)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1044
517	29213416 (15051, 15052)	Novel Protein sim. GBank gil1341021spl P08823 RUBA_WHEAT - RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT PRECURSOR (60 KD CHAPERONIN ALPHA SUBUNIT) (CPN-60 ALPHA)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1044
518	30256541 (3501, 3502)	Novel Protein sim. GBank gil1587206 pf 2206327A - T complex protein [Cucumis sativus]	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1024
519	80237325 (14191, 14192)	Novel Protein sim. GBank gil1729867 spl P12612 TCPA_YEAST - T-COMPLEX PROTEIN 1, ALPHA SUBUNIT (TCP-1-ALPHA) (CCT-ALPHA)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1004, 1006, 1017, 1023, 1034
520	10206946 (6959, 6960)	Novel Protein sim. GBank gil1899190 (U90204) - heat shock protein 60 [Tsukamurella tyrosinolvens]	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1003
521	86475368 (7441, 7442)	Novel Protein sim. GBank gil1899190 (U90204) - heat shock protein 60 [Tsukamurella tyrosinolvens]	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1027, 1048
522	85800720 (13739, 13740)	Novel Protein sim. GBank gil1899190 (U90204) - heat shock protein 60 [Tsukamurella tyrosinolvens]	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1025
523	79471293 (16029, 16030)	Novel Protein sim. GBank gil231752 spl Q00767 CH61_STRAL - 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1) (HSP58)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1008, 1022, 1026, 1029

524	80216600 (5031, 5032)	Novel Protein sim. GBank gi 462047 sp P34756 FAB1_YEAST - PROBABLE PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE FAB1 (1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE KINASE) (PIP5K) (PTDINS(4)P-5-KINASE) (DIPHOSPHOINOSITIDE KINASE)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1001, 1006, 1008, 1010, 1022, 1024, 1026, 1038, 1046
525	8345187 (22281, 22282)	Novel Protein sim. GBank gi 493574 (U09480) - Bin2p [Saccharomyces cerevisiae]	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1024
526	80221080 (17033, 17034)	Novel Protein sim. GBank gi 730923 sp P39079 TCPZ_YEAST - T-COMPLEX PROTEIN 1, ZETA SUBUNIT (TCP-1-ZETA) (CCT-ZETA)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1001, 1004, 1006, 1008, 1010, 1024, 1025, 1031
527	80236306 (22441, 22442)	Novel Protein sim. GBank gi 730923 sp P39079 TCPZ_YEAST - T-COMPLEX PROTEIN 1, ZETA SUBUNIT (TCP-1-ZETA) (CCT-ZETA)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	1003, 1006, 1017, 1025
528	20294481 (18905, 18906)	Novel Protein sim. GBank gi 123677 sp P02829 HS82_YEAST - HEAT SHOCK PROTEIN HSP82	Contains protein domain (PF001183) - Hsp90 protein	eph	1034
529	798377480 (7199, 7200)	Novel Protein sim. GBank gi 123726 sp P10413 HTPG_ECOLI - HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G) (HEAT SHOCK PROTEIN C62.5)	Contains protein domain (PF001183) - Hsp90 protein	eph	1001, 1008
530	378005887 (7965, 7966)	Novel Protein sim. GBank gi 123726 sp P10413 HTPG_ECOLI - HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G) (HEAT SHOCK PROTEIN C62.5)	Contains protein domain (PF001183) - Hsp90 protein	eph	1012
531	80427330 (7863, 7864)	Novel Protein sim. GBank gi 417154 sp P33126 HS82_ORYSA - HEAT SHOCK PROTEIN 82	Contains protein domain (PF001183) - Hsp90 protein	eph	1025, 1031, 1050
532	11304133 (8301, 8302)	Novel Protein sim. GBank gi 284249 emb CAA16887.1 - (AL021749) heat-shock protein [Arabidopsis thaliana]	Contains protein domain (PF00226) - DnaJ domain	eph	1024

533	80504390 (10229, 10230)	Novel Protein sim. GBank gi 42118137 emb CAA10745I - (AJ132708) Dnaj1 protein [Anabaena sp.]	Contains protein domain (PF00226) - Dnaj domain	eph	1012
534	80021335 (5219, 5220)	Novel Protein sim. GBank gi 731164 sp P39102 XDJ1_YEAST - XDJ1 PROTEIN	Contains protein domain (PF00226) - Dnaj domain	eph	1022, 1044
535	79856214 (9725, 9726)	Novel Protein sim. GBank gi 125321 sp P00556 KDGL_ECOLI - DIACYLGLYCEROL KINASE (DAGK) (DIGLYCERIDE KINASE) (DGK)	Contains protein domain (PF01219) - Prokaryotic diacylglycerol kinase	eph	1017
536	39704657 (4215, 4216)	Novel Protein sim. GBank gi 1076746 pir S53126 - heat shock protein 70 - rice (fragment)		eph	1017, 1044
537	80052618 (475, 476)	Novel Protein sim. GBank gi 1170730 sp P44446 LCFH_HAEIN - PUTATIVE LONG-CHAIN-FATTY-ACID-COA LIGASE (LONG-CHAIN ACYL-COA SYNTHETASE) (LACS)		eph	1010, 1034, 1050
538	78463277 (4275, 4276)	Novel Protein sim. GBank gi 123593 sp P11143 HS70_MAIZE - HEAT SHOCK 70 KD PROTEIN		eph	1008
539	94639561 (11639, 11640)	Novel Protein sim. GBank gi 123634 sp P22202 HS74_YEAST - HEAT SHOCK PROTEIN SSA4		eph	1001, 1003, 1004, 1008, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1026, 1027, 1029, 1031, 1034, 1038, 1039, 1044, 1054
540	57293325 (14465, 14466)	Novel Protein sim. GBank gi 126011 sp P18163 LCFB_RAT - LONG-CHAIN- FATTY-ACID-COA LIGASE, LIVER ISOZYME (LONG-CHAIN ACYL-COA SYNTHETASE 2) (LACS 2)		eph	1022

541	56148637 (15685, 15686)	Novel Protein sim. GBank gi 1651678 dbj BAA16606 - (D90899) hypothetical protein [Synechocystis sp.]		eph	1003
542	87899128 (1825, 1826)	Novel Protein sim. GBank gi 1708467 sp P50019 DNAK_AGRTU - DNAK PROTEIN (HEAT SHOCK PROTEIN 70) (HSP70)		eph	1015
543	65494578 (2031, 2032)	Novel Protein sim. GBank gi 1808585 emb CAA67132 - (X98508) heat shock protein 70 [Agaricus bisporus]		eph	1023
544	86474044 (19969, 19970)	Novel Protein sim. GBank gi 1931652 U95573 - phosphatidylinositol-4-phosphate 5-kinase isolog [Arabidopsis thaliana]		eph	1023, 1054
545	57568071 (22697, 22698)	Novel Protein sim. GBank gi 2127959 pir H64509 - heat shock protein X - Methanococcus jannaschii		eph	1024, 1034
546	11011944 (11567, 11568)	Novel Protein sim. GBank gi 2145715 pir IS72599_B1937_F2_56 protein - Mycobacterium leprae		eph	1027
547	27965593 (11011, 11012)	Novel Protein sim. GBank gi 2314540 gb AAD08413.1 - (AE000637) type III restriction enzyme M protein (mod) [Helicobacter pylori 26965]		eph	1006
548	79224056 (18041, 18042)	Novel Protein sim. GBank gi 2309505 emb CAA17390 - (AL021931) clpB [Mycobacterium tuberculosis]		eph	1024
549	80235193 (12541, 12542)	Novel Protein sim. GBank gi 2380281 emb CAA50548 - (X71416) TOR2 [Saccharomyces cerevisiae]		eph	1006, 1025, 1026
550	79558866 (15117, 15118)	Novel Protein sim. GBank gi 2380281 emb CAA50548 - (X71416) TOR2 [Saccharomyces cerevisiae]		eph	1025
551	20613117 (21323, 21324)	Novel Protein sim. GBank gi 3061312 dbj BAA25664 - (AB006916) phosphatidylinositol-4-phosphate 5-kinase I-gamma [Mus musculus]		eph	1022

552	57301161 (6917, 6918)	Novel Protein sim. GBank gi 3493666 (AF085219) - alpha dacylglycerol kinase; alpha-DGK [Mus musculus]		eph	1006
553	9499993 (16923, 16924)	Novel Protein sim. GBank gi 3914345 sp O70172 PI52_MOUSE - PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE TYPE II ALPHA (PIP5KII-ALPHA) (1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE KINASE) (PTDINS(4)P-5-KINASE B ISOFORM) (DIPHOSPHOINOSITIDE KINASE)		eph	1022, 1034
554	11398513 (21289, 21290)	Novel Protein sim. GBank gi 4001713 dbj BA-A35087.1 - (AB019879) DnaK [Porphyromonas gingivalis]		eph	1006
555	20711449 (1207, 1208)	Novel Protein sim. GBank gi 4323240 gb AAD16273.1 - (AF098972) phosphatidylinositol 3-kinase TOR1 [Filobasidiella neoformans]		eph	1006
556	8498616 (13147, 13148)	Novel Protein sim. GBank gi 462047 sp P34756 FAB1_YEAST - PROBABLE PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE FAB1 (1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE KINASE) (PIP5K) (PTDINS(4)P-5-KINASE) (DIPHOSPHOINOSITIDE KINASE)		eph	1034
557	10077209 (21847, 21848)	Novel Protein sim. GBank gi 475385 embj CAB42048.1 - (AL049754) putative ATP-dependent protease ATP-binding subunit [Streptomyces coelicolor]		eph	1038
558	86380106 (20667, 20668)	Novel Protein sim. GBank gi 4759142 ref NP_004776.1 pSLC9 - solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 2		eph	1053
559	25260575 (6913, 6914)	Novel Protein sim. GBank gi 545382 ref NP_006380.1 pORP1 - oxygen regulated protein (150kD)		eph	1008

560	78748286 (16271, 16272)	Novel Protein sim. GBank gil586048 sp P37297 STT4_YEAST - PHOSPHATIDYLINOSITOL 4-KINASE STT4 (PI4-KINASE) (PTDINS-4-KINASE)		eph	1044
561	30271154 (7681, 7682)	Novel Protein sim. GBank gil729925 sp P39002 LCF3_YEAST - LONG-CHAIN-FATTY-ACID-COA LIGASE 3 (LONG-CHAIN ACYL-COA SYNTHETASE 3) (FATTY ACID ACTIVATOR 3)		eph	1026
562	706337463 (17057, 17058)	Novel Protein sim. GBank gil95711 pir D35905 - ATP-dependent Clp proteinase (EC 3.4.21.-) regulatory chain B - Escherichia coli		eph	1049
563	80061643 (21127, 21128)	Novel Protein sim. GBank gil149527 emb CAA96487 - (Z71928) para-nitrobenzyl esterase [Bacillus subtilis]	Contains protein domain (PF00135) - Carboxylesterases	esterase	1024
564	16844643 (2211, 2212)	Novel Protein sim. GBank gil2896741 emb CAA17220.1 - (AL021897) hypothetical protein Rv1104 [Mycobacterium tuberculosis]	Contains protein domain (PF00135) - Carboxylesterases	esterase	1023
565	80248149 (14945, 14946)	Novel Protein sim. GBank gil295642 (L13036) - phospholipase C [Saccharomyces cerevisiae]	Contains protein domain (PF00388) - Phosphatidylinositol-specific phospholipase C, X domain	esterase	1003, 1006, 1017, 1022
566	38916455 (3215, 3216)	Novel Protein sim. GBank gil135612 sp P23911 TESB_ECOLI - ACYL-COA THIOESTERASE II		esterase	1026
567	80248144 (21125, 21126)	Novel Protein sim. GBank gil135612 sp P23911 TESB_ECOLI - ACYL-COA THIOESTERASE II		esterase	1001, 1006, 1024, 1026
568	80245040 (21307, 21308)	Novel Protein sim. GBank gil1705954 sp P22434 CNA1_YEAST - 3',5'-CYCLIC-NUCLEOTIDE PHOSPHODIESTERASE 1 (PDEASE 1) (LOW-AFFINITY CAMP PHOSPHODIESTERASE)		esterase	1001, 1003, 1004, 1006, 1008, 1010, 1017, 1022, 1023, 1024, 1025, 1027, 1034, 1039, 1044

569	80243935 (21989, 21990)	Novel Protein sim. GBank gli172229 (M17781) - low-affinity cyclic AMP phosphodiesterase [Saccharomyces cerevisiae]		esterase	1001, 1003, 1004, 1006, 1008, 1010, 1017, 1022, 1024, 1034, 1038, 1044, 1054
570	79559694 (11709, 11710)	Novel Protein sim. GBank gli2506993 sp P41407 ACPD_ECOLI - ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE)		esterase	1038
571	80086554 (14951, 14952)	Novel Protein sim. GBank gli2982501 emb CAA061641 - (AJ004832) neuropathy target esterase [Homo sapiens]		esterase	1022, 1024
572	87898870 (16541, 16542)	Novel Protein sim. GBank gli3282794 (AF04495) - CheB homolog [Agrobacterium tumefaciens]		esterase	1015
573	79574481 (10705, 10706)	Novel Protein sim. GBank gli348403 pir A44832 - esterase estA - Pseudomonas sp		esterase	1031
574	39457148 (6179, 6180)	Novel Protein sim. GBank gli3649751 emb CAA78842] - (Z15137) esterase A [Streptomyces chrysomallus]		esterase	1024
575	79830591 (6287, 6288)	Novel Protein sim. GBank gli4808354 emb CAB42768_1] - (AL049841) putative esterase [Streptomyces coelicolor]		esterase	1039
576	25136426 (19977, 19978)	Novel Protein sim. GBank gli7319361 sp P40363 YJG8 - HYPOTHETICAL 33.9 KD ESTERASE IN SMC3-MRPL8 INTERGENIC REGION		esterase	1026
577	80243851 (9387, 9388)			esterase	1003, 1004, 1034

578	17939082 (18693, 18694)	Novel Protein sim. GBank gi 1705460 sp P53555 BIOA_BACSU - ADENOSYL METHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7,8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate	gaba	1039
579	87451065 (18167, 18168)	Novel Protein sim. GBank gi 2829813 sp P9490 ARGD_MYCTU - ACETYLORNITHINE AMINOTRANSFERASE (ACOAT)	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate	gaba	1017, 1018
580	15033376 (1255, 1256)	Novel Protein sim. GBank gi 3395432 (AC004683) - unknown protein [Arabidopsis thaliana]	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate	gaba	1054
581	29685126 (2795, 2796)	Novel Protein sim. GBank gi 4151931 gb AA09412.1 - (AF110737) RhbA [Sinorhizobium meliloti]	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate	gaba	1009, 1024
582	15024413 (8185, 8186)	Novel Protein sim. GBank gi 135040 sp P07768 SUIS_RABIT - SUCRASE-ISOMALTASE, INTESTINAL /	Contains protein domain (PF01055) - Glycosyl hydrolases family 31	glucoamylase	1054
583	80586583 (89, 90)	Novel Protein sim. GBank gi 2497790 sp P70699 LYAG_MOUSE - LYSOSOMAL ALPHA-GLUCOSIDASE PRECURSOR (ACID MALTASE)	Contains protein domain (PF01055) - Glycosyl hydrolases family 31	glucoamylase	1004, 1006, 1010, 1024, 1058
584	46882290 (5053, 5054)	Novel Protein sim. GBank gi 4758712 ref NP_004659.1 pMG - UNKNOWN	Contains protein domain (PF01055) - Glycosyl hydrolases family 31	glucoamylase	1029
585	8754180 (22789, 22790)	Novel Protein sim. GBank gi 172525 (M16165) - S2 protein [Saccharomyces cerevisiae]		glucoamylase	1034
586	80106002 (4055, 4056)	Novel Protein sim. GBank gi 552087 (M33753) - crumbs protein [Drosophila melanogaster]	Contains protein domain (PF00008) - EGF-like domain	glycoprotein	1016, 1053, 1054

587	27837511 (16185, 16186)	Novel Protein sim. GBank g 17098 3 sp P80667 PEXD_YEAST - PEROXISOMAL MEMBRANE PROTEIN PAS20 (PEROXIN-13)	Contains protein domain (PF00018) - SH3 domain	glycoprotein	1010, 1022, 1024, 1034, 1054
588	78936486 (22301, 22302)	Novel Protein sim. GBank g 4504489 ref NP_000403.1 pHRG - histidine-rich glycoprotein	Contains protein domain (PF00031) - Cy/statin domain	glycoprotein	1008, 1022, 1026, 1039, 1044
589	55307909 (2107, 2108)	Novel Protein sim. GBank g 185502 L02325 - Immunoglobulin heavy chain [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	glycoprotein	1019
590	80048814 (22275, 22276)	Novel Protein sim. GBank g 24973 1 sp P5583 MOG_BOVIN - MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR	Contains protein domain (PF00047) - Immunoglobulin domain	glycoprotein	1008, 1010, 1030, 1041
591	25338668 (8789, 8790)	Novel Protein sim. GBank g 1083829 pir1 XK0353 - zinc-alpha 2-glycoprotein - rat PROTEIN	Contains protein domain (PF00129) - Class I Histocompatibility antigen, domains alpha 1 and 2	glycoprotein	1026
592	13058280 (11991, 11992)	Novel Protein sim. GBank g 46261 5 sp P21590 MRP_ECOLI - MRP PROTEIN	Contains protein domain (PF00142) - 4Fe-4S iron sulfur cluster binding proteins, NifH/frcC family	glycoprotein	1031
593	950011920 (10117, 10118)	Novel Protein sim. GBank g 113560 sp P02770 ALBU_RAT - SERUM ALBUMIN PRECURSOR	Contains protein domain (PF00273) - Serum albumin family	glycoprotein	1006, 1022, 1039
594	79483947 (7931, 7932)	Novel Protein sim. GBank g 548938 sp Q06855 SLP2_CLOTM - CELL SURFACE GLYCOPROTEIN 2 PRECURSOR (S-LAYER PROTEIN 2)	Contains protein domain (PF00395) - S-layer homology domain containing proteins	glycoprotein	1006
595	16532807 (17519, 17520)	Novel Protein sim. GBank g 2135586 pir1 38004 - M130 antigen (cytosolic variant 1) - human	Contains protein domain (PF000530) - Scavenger receptor cysteine-rich domain	glycoprotein	1030
596	17893306 (4585, 4586)	Novel Protein sim. GBank g 3183011 sp O08770 GPV_RAT - PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D)	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	1008

597 14894)	79592218 (14893, Novel Protein sim. GBank g 4206629 (AF067954) [Salmonella typhimurium] - putative membrane fusion protein SII/B	Contains protein domain (PF00873) - AcrB/AcrD/AcrF family	glycoprotein	1024
598 3088)	78779991 (3087, Novel Protein sim. GBank g 1723794 sp P53198 ERP6 _ YEAST - ERP6 PROTEIN PRECURSOR	Contains protein domain (PF01105) - emp24/gp25/jp24 family	glycoprotein	1008
599 11922)	80234323 (11921, Novel Protein sim. GBank g 1175955 sp P43553 ALR2 _ YEAST - ALUMINUM RESISTANCE PROTEIN 2	Contains protein domain (PF01544) - CorA-like Mg2+ transporter protein	glycoprotein	1003, 1004, 1006, 1008, 1024, 1026, 1038
600 15410)	38433999 (15409, Novel Protein sim. GBank g 124953 sp P24063 ITAL _ MOUSE - LEUKOCYTE ADHESION GLYCOPROTEIN LFA- 1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (INTEGRIN ALPHA-L)	Contains protein domain (PF01839) - FG-GAP repeat	glycoprotein	1008
601 16562)	66259053 (16561, Novel Protein sim. GBank g 1030956 (U26160) - 43 kDa secreted glycoprotein precursor [Paracoccidioides brasiliensis]		glycoprotein	1010
602 9800)	78385717 (9799, Novel Protein sim. GBank g 11503 pir S15674 - cell surface glycoprotein OX47 precursor - rat		glycoprotein	1003
603 16422)	86667119 (16421, Novel Protein sim. GBank g 122223 emb CA63862 - (X94148) putative inner membrane protein [Bacillus licheniformis]		glycoprotein	1029
604 12994)	20622174 (12993, Novel Protein sim. GBank g 12831 sp P25754 601M_PSEPU - 60 KD INNER-MEMBRANE PROTEIN		glycoprotein	1004
605 9280)	77722368 (9279, Novel Protein sim. GBank g 113668 sp P23961 ALLC_HUMAN - !!! ALU CLASS C WARNING ENTRY !!!		glycoprotein	1054
606 18930)	20273344 (18929, Novel Protein sim. GBank g 113750 sp P15684 AMPN_RAT - AMINOPEPTIDASE N (MICROSOMAL AMINOPEPTIDASE)		glycoprotein	1022

607	66376915 (16983, 16984)	Novel Protein sim. GBank gj 114113 spl P13981 ARCA_PSEAE - ARGININE DEIMINASE		glycoprotein	1010
608	10225796 (1711, 1712)	Novel Protein sim. GBank gj 1176192 spl P45420 YHCD_ECOLI - HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN IN GLTF-NANT INTERGENIC REGION PRECURSOR		glycoprotein	1039
609	71811433 (13487, 13488)	Novel Protein sim. GBank gj 1209711 (U41293) - unknown [Saccharomyces cerevisiae]		glycoprotein	1008
610	84428046 (41, 42)	Novel Protein sim. GBank gj 1224131 (U50403) - breast cancer suppressor element Ishmael Upper RP2 [Homo sapiens]		glycoprotein	1044, 1053
611	80306250 (1741, 1742)	Novel Protein sim. GBank gj 123530 spl P04929 HRPX_PLAO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR		glycoprotein	1006, 1024, 1025, 1029
612	17882406 (2173, 2174)	Novel Protein sim. GBank gj 124905 spl P27772 IRGA_VIBCH - IRON-REGULATED OUTER MEMBRANE VIRULENCE PROTEIN PRECURSOR		glycoprotein	1039
613	86381114 (8851, 8852)	Novel Protein sim. GBank gj 126377 spl P11438 LAMP1_MOUSE - LYSOSOME-ASSOCIATED MEMBRANE GLYCOPROTEIN 1 PRECURSOR (LAMP-1) (LGP-A) (LGP-120) (CD107A) (P2B)		glycoprotein	1023, 1053
614	78675350 (18349, 18350)	Novel Protein sim. GBank gj 131385 spl P26320 PSBO_SOLTU - OXYGEN-EVOLVING ENHANCER PROTEIN 1 PRECURSOR (OEE1) (33 KD SUBUNIT OF OXYGEN EVOLVING SYSTEM OF PHOTOSYSTEM II) (33 KD THYLAKOID MEMBRANE PROTEIN)		glycoprotein	1009
615	80499327 (14707, 14708)	Novel Protein sim. GBank gj 1314835 (U56084) - exogenous ferric siderophore receptor [Bordetella bronchiseptica]		glycoprotein	1012

616	20385684 (16451, 16452)	Novel Protein sim. GBank g1133321spl P16356 RPB1, CAAEEL - DNA- DIRECTED RNA POLYMERASE II LARGEST SUBUNIT		glycoprotein	1004
617	11072703 (779, 780)	Novel Protein sim. GBank g11344011spl P19674 SECF_ECOLI - PROTEIN- EXPORT MEMBRANE PROTEIN SECF		glycoprotein	1034
618	79965570 (1149, 1150)	Novel Protein sim. GBank g11352941spl P47179 Y9P_YEAST - HYPOTHETICAL 118.4 KD PROTEIN IN BAT2- DAL5 INTERGENIC REGION PRECURSOR		glycoprotein	1009, 1010
619	79829688 (10525, 10526)	Novel Protein sim. GBank g11364094 pirl S57509 - Integral membrane protein - Streptomyces pristinaespiralis		glycoprotein	1003
620	79242696 (21883, 21884)	Novel Protein sim. GBank g11404121spl P25594 YCG9_YEAST - HYPOTHETICAL 50.2 KD PROTEIN IN HML 5'REGION		glycoprotein	1008, 1022
621	32153098 (6771, 6772)	Novel Protein sim. GBank g11498192 dbj BAA05665 - (D28119) putative [Pseudomonas aeruginosa]		glycoprotein	1031
622	11698153 (6943, 6944)	Novel Protein sim. GBank g11504107 emb CAB01714 - (Z78535) outer membrane usher protein [Proteus mirabilis]		glycoprotein	1039
623	28782535 (8225, 8226)	Novel Protein sim. GBank g1165166f dbj BAA16594 - (D90899) ferric aerobactin receptor [Synechocystis sp.]		glycoprotein	1022
624	71768088 (12851, 12852)	Novel Protein sim. GBank g11652538 dbj BAA17489 - (D90906) nitrate transport protein NrtD [Synechocystis sp.]		glycoprotein	1026
625	25262451 (21035, 21036)	Novel Protein sim. GBank g11703287 spl Q11010 AMPN_STRLI - AMINOPEPTIDASE N (LYSYL AMINOPEPTIDASE) (LYS-AP) (ALANINE AMINOPEPTIDASE)		glycoprotein	1008

626	54994452 (9161, 9162)	Novel Protein sim. GBank gii 1708794 sp P37949 LEPA_BACSU - GTP-BINDING PROTEIN LEPA		glycoprotein	1029
627	94669340 (177803, 17804)	Novel Protein sim. GBank gii 1730615 sp P54837 ERV5_YEAST - ERV25 PROTEIN PRECURSOR		glycoprotein	1004, 1010, 1023
628	77513249 (13561, 13562)	Novel Protein sim. GBank gii 1778500 (U82598) - ferrierobactin receptor precursor [Escherichia coli]		glycoprotein	1009, 1023, 1031, 1049
629	80256055 (17783, 17784)	Novel Protein sim. GBank gii 1787068 (AE000186) - putative DEOR-type transcriptional regulator [Escherichia coli]		glycoprotein	1034
630	1008603 (5101, 5102)	Novel Protein sim. GBank gii 1788678 (AE000322) - putative outer membrane protein [Escherichia coli]		glycoprotein	1017
631	7520262 (20691, 20692)	Novel Protein sim. GBank gii 1816664 (U86522) - platelet membrane glycoprotein Ib (GPIb) specific antibody [Homo sapiens]		glycoprotein	1058
632	79617790 (40445, 4046)	Novel Protein sim. GBank gii 2121002 pir 40333 - tracheal colonization factor - Bordetella pertussis		glycoprotein	1003
633	79491185 (20443, 20444)	Novel Protein sim. GBank gii 2129478 pir S51939 - chitinase (EC 3.2.1.14); precursor - beet		glycoprotein	1050
634	9999167 (3921, 3922)	Novel Protein sim. GBank gii 2131317 pir S67570 - hypothetical protein YDL037c - yeast (Saccharomyces cerevisiae)		glycoprotein	1039
635	7925651 (17427, 17428)	Novel Protein sim. GBank gii 2132268 pir S69069 - hypothetical protein YPR083w - yeast (Saccharomyces cerevisiae)		glycoprotein	1003
636	78985527 (12183, 12184)	Novel Protein sim. GBank gii 2132377 pir S69459 - hypothetical protein YPR197c - yeast (Saccharomyces cerevisiae)		glycoprotein	1017
637	77803302 (13188, 13190)	Novel Protein sim. GBank gii 227029 pir 16 3275A - romA gene [Enterobacter cloacae]		glycoprotein	1051

638 12448)	21641132 (12447, Novel Protein sim. GBank gi 231468 sp P24090 A2HS_RAT - ALPHA-2-HS- GLYCOPROTEIN PRECURSOR (FETUIN) (GLYCOPROTEIN PP63) (59 KD BONE SIALIC ACID-CONTAINING PROTEIN) (BSP)		glycoprotein	1004, 1022
639 6126)	95290209 (6725, Novel Protein sim. GBank gi 2425172 dbj BAAA22267.1 - (AB006797) OprE3 [Pseudomonas aeruginosa]		glycoprotein	1004
640 12938)	9383468 (12937, Novel Protein sim. GBank gi 2425172 dbj BAAA22267.1 - (AB006797) OprE3 [Pseudomonas aeruginosa]		glycoprotein	1017
641 7996)	87456786 (7995, Novel Protein sim. GBank gi 2462748 (AC002292) - putative Clathrin Coat Assembly protein [Arabidopsis thaliana]		glycoprotein	1025
642 19124)	27356100 (19123, Novel Protein sim. GBank gi 2492680 sp Q999186 APM4_YEAST - ADAPTIN MEDIUM CHAIN HOMOLOG APM4		glycoprotein	1026
643 180)	27835570 (179, Novel Protein sim. GBank gi 2498568 sp Q48899 MP1_MYCAV - MAJOR MEMBRANE PROTEIN I (MP1) (35 KD ANTIGEN)		glycoprotein	1006
644 8054)	80218455 (8053, Novel Protein sim. GBank gi 2498760 sp Q12462 PEXB_YEAST - PEROXISOMAL MEMBRANE PROTEIN PMP27 (PEROXIN-11)		glycoprotein	1006, 1008, 1009, 1010, 1022, 1024, 1025
645 8824)	20703281 (8823, Novel Protein sim. GBank gi 2507398 sp P23931 MLTD_ECOLI - MEMBRANE-BOUND LYtic MUREIN TRANSGLYCOSYLASE D PRECURSOR (MUREIN HYDROLASE D) (REGULATORY PROTEIN DNIR)		glycoprotein	1010
646 22122)	78527678 (22121, Novel Protein sim. GBank gi 2507462 sp P17315 CIRA_ECOLI - COLICIN I RECEPTOR PRECURSOR		glycoprotein	1003
647 10398)	28475335 (10397, Novel Protein sim. GBank gi 2618486 dbj BAAA23372 - (AB008553) mLGP85/LIMP II [Mus musculus]		glycoprotein	1026

648	78376175 (7181, 7182)	Novel Protein sim. GBank gil2649154 (AE001006) - membrane protein [Archaeoglobus fulgidus]		glycoprotein	1026
649	65665626 (17281, 17282)	Novel Protein sim. GBank gil2649577 (AE001033) - conserved hypothetical protein [Archaeoglobus fulgidus]		glycoprotein	1016
650	94240170 (6071, 6072)	Novel Protein sim. GBank gil2739028 (AF023919) - PK-120 precursor [Mus musculus]		glycoprotein	1017, 1039, 1044
651	19877114 (979, 980)	Novel Protein sim. GBank gil2808460 emb CAA75431] - (Y15162) ptk [Acinetobacter johnsonii]		glycoprotein	1034
652	20438295 (15721, 15722)	Novel Protein sim. GBank gil2808721 emb CAA16245] - (AL021428) hypothetical protein Rv0064 [Mycobacterium tuberculosis]		glycoprotein	1010
653	39565023 (17871, 17872)	Novel Protein sim. GBank gil2829634 isp P75857 YCBS_ECOLI - HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN IN PEPN-PYRD INTERGENIC REGION PRECURSOR		glycoprotein	1004
654	78461524 (4095, 4096)	Novel Protein sim. GBank gil2909549 emb CAA17428] - (AL021933) hypothetical protein Rv0473 [Mycobacterium tuberculosis]		glycoprotein	1008, 1026
655	30369344 (6101, 6102)	Novel Protein sim. GBank gil2911094 emb CAA17478] - (AL021957) hypothetical protein Rv2174 [Mycobacterium tuberculosis]		glycoprotein	1054
656	80501805 (22477, 22478)	Novel Protein sim. GBank gil2959367 emb CAA17921] - (AL022117) hypothetical protein [Schizosaccharomyces pombe]		glycoprotein	1012, 1023, 1044
657	25143546 (6501, 6502)	Novel Protein sim. GBank gil2983432 (AE000773) - erythrocyte band 7 homolog [Aquifex aeolicus]		glycoprotein	1008

658	8516846 (6087, 6088)	Novel Protein sim. G-Bank gi 3006160 emb CAA18420 - (AL022305) mitochondrial import inner membrane translocase subunit precursor [Schizosaccharomyces pombe]		glycoprotein	1034
659	56687719 (3211, 3212)	Novel Protein sim. G-Bank gi 3024882 emb O05916 Y1B1_MYCTU - HYPOTHETICAL 45.0 KD PROTEIN CY21C12.18C		glycoprotein	1006, 1009
660	78378781 (19341, 19342)	Novel Protein sim. G-Bank gi 3025119 emb O05518 YDIE_BACSU - HYPOTHETICAL 36.8 KD PROTEIN IN PHOB- GROES INTERGENIC REGION		glycoprotein	1008, 1026
661	55777288 (21565, 21566)	Novel Protein sim. G-Bank gi 3080474 emb CAA18669 - (AL022602) cell division protein FisW [Mycobacterium leprae]		glycoprotein	1042
662	86474053 (22277, 22278)	Novel Protein sim. G-Bank gi 3212871 (AC004005) - putative translation initiation factor [Arabidopsis thaliana]		glycoprotein	1054
663	91012401 (22557, 22558)	Novel Protein sim. G-Bank gi 3261795 emb CA09497 - (Z96072) hypothetical protein Rv2690c [Mycobacterium tuberculosis]		glycoprotein	1010
664	79603142 (9423, 9424)	Novel Protein sim. G-Bank gi 3261829 emb CA09271 - (Z98260) hypothetical protein Rv1230c [Mycobacterium tuberculosis]		glycoprotein	1003, 1039
665	79262079 (22388, 22390)	Novel Protein sim. G-Bank gi 3413391 emb CAA20252 - (AL031231) hypothetical protein SC3C3.03c [Streptomyces coelicolor]		glycoprotein	1008
666	78077198 (18829, 18830)	Novel Protein sim. G-Bank gi 3413446 emb CAA72351 - (Y11648) wlaB [Campylobacter jejuni]		glycoprotein	1029

667	27975500 (12393, 12394)	Novel Protein sim. GBank gi 3860837 emb CAA14737 - (AJ235271) GTP-BINDING PROTEIN LEPA (lepa) [Rickettsia prowazekii]		glycoprotein	1022
668	11094066 (21487, 21488)	Novel Protein sim. GBank gi 3860837 emb CAA14737 - (AJ235271) GTP-BINDING PROTEIN LEPA (lepa) [Rickettsia prowazekii]		glycoprotein	1022
669	80105792 (1387, 1388)	Novel Protein sim. GBank gi 3913647 spi Q6155 FBN2_MOUSE - FIBRILLIN 2 PRECURSOR		glycoprotein	1013, 1024
670	80087095 (21157, 21158)	Novel Protein sim. GBank gi 3915963 spi P32386 YBT1_YEAST - ATP-DEPENDENT BILE ACID PERMEASE		glycoprotein	1008, 1022, 1038
671	17729486 (7991, 7992)	Novel Protein sim. GBank gi 3941269 (AF018093) - similarity to SCAMP37 [Pisum sativum]		glycoprotein	1003
672	29004469 (8525, 8526)	Novel Protein sim. GBank gi 39942 spi P24180 ACRE_ECOLI - ACRIFLAVIN RESISTANCE PROTEIN E PRECURSOR (ENVCP PROTEIN)		glycoprotein	1003
673	20285913 (19211, 19212)	Novel Protein sim. GBank gi 400767 emb CAA22357 - (AL034443) putative integral membrane efflux protein [Streptomyces coelicolor]		glycoprotein	1017, 1034
674	11249975 (18669, 18670)	Novel Protein sim. GBank gi 408092 (L24961) - ORF; Putative [Saccharomyces cerevisiae]		glycoprotein	1010
675	32168560 (4399, 4400)	Novel Protein sim. GBank gi 41751 spi P32272 PORD_PSEAE - PORIN D PRECURSOR (OUTER MEMBRANE PROTEIN D2) (IMIPENEM/BASIC AMINO ACID-SPECIFIC OUTER MEMBRANE PORE)		glycoprotein	1023
676	78751013 (20699, 20700)	Novel Protein sim. GBank gi 418265 spi P3234 VM12_YEAST - VACUOLAR ATPASE ASSEMBLY INTEGRAL MEMBRANE PROTEIN VMA12 (PROTEIN VPH2)		glycoprotein	1044

677	20288211 (13581, 13582)	Novel Protein sim. GBank gi 435039 (L17308) - proline-rich cell wall protein [Gossypium hirsutum]		glycoprotein	1022
678	25245755 (8363, 8364)	Novel Protein sim. GBank gi 4589961 gb AAD26479.1 AC00716 - (AC007169) unknown protein [Arabidopsis thaliana]		glycoprotein	1023
679	8910018 (2655, 2656)	Novel Protein sim. GBank gi 479628 pir S34969 - parin E1 - Pseudomonas aeruginosa		glycoprotein	1008
680	80248729 (17043, 17044)	Novel Protein sim. GBank gi 4808349 emb CA842763.1 - (AL049841) putative membrane protein [Streptomyces coelicolor]		glycoprotein	1006, 1008, 1023, 1026, 1029, 1039, 1044
681	9887562 (15255, 15256)	Novel Protein sim. GBank gi 5091482 gb AAD39555.1 - (AF031417) PhaK- like protein [Pseudomonas putida]		glycoprotein	1034
682	95292928 (22323, 22324)	Novel Protein sim. GBank gi 5102807 emb CA845222.1 - (AL079308) putative membrane protein [Streptomyces coelicolor]		glycoprotein	1024
683	79199080 (21131, 21132)	Novel Protein sim. GBank gi 511500 (L29116) - Ig rearranged mu-chain VH4-DXP1-JH5 region [Homo sapiens]		glycoprotein	1024
684	80504392 (10547, 10548)	Novel Protein sim. GBank gi 5354138 gb AAD42397.1 AF15749 - (AF157493) conserved hypothetical integral membrane protein [Zymomonas mobilis]		glycoprotein	1009, 1010, 1012, 1044, 1049
685	42563257 (22381, 22382)	Novel Protein sim. GBank gi 5454098 ref NP_06465.1 pT1A - lung type-I cell membrane-associated glycoprotein		glycoprotein	1001
686	80588678 (7219, 7220)	Novel Protein sim. GBank gi 549651 sp P36172 YK85 - YEAST - HYPOTHETICAL 63.4 KD PROTEIN IN SIR1 3'REGION		glycoprotein	1022, 1058

687	866687071 (8903, 8904)	Novel Protein sim. G-Bank gil553847 (M18193) - inter-alpha-trypsin inhibitor heavy chain old gene name 'T1' [Homo sapiens]		glycoprotein	1029
688	80050121 (21659, 21660)	Novel Protein sim. G-Bank gil5670176 gb AAAD46616_1 AF16131 - (AF161317) NRAMP manganese transport protein Mrta [Salmonella typhimurium]		glycoprotein	1004, 1008, 1034
689	37805958 (6355, 6356)	Novel Protein sim. G-Bank gil586420 sp P38065 YBD7 _ YEAST - ALPHA- ADAPTIN HOMOLOG IN URA7-POL12 INTERGENIC REGION		glycoprotein	1058
690	54856600 (4405, 4406)	Novel Protein sim. G-Bank gil72883-1 sp P3918 ALU1_HUMAN - !!! ALU SUBFAMILY J WARNING ENTRY !!!		glycoprotein	1019
691	37802938 (411, 412)	Novel Protein sim. G-Bank gil729835 sp P40190 IL6B_RAT - INTERLEUKIN- 6 RECEPTOR BETA CHAIN PRECURSOR (IL-6R BETA) (INTERLEUKIN 6 SIGNAL TRANSDUCER) (MEMBRANE GLYCOPROTEIN 130) (GP130)		glycoprotein	1058
692	78749244 (2135, 2136)	Novel Protein sim. G-Bank gil730249 sp P39685 P152_YEAST - NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM152 (P150)		glycoprotein	1022, 1026
693	9841963 (14039, 14040)	Novel Protein sim. G-Bank gil78921 pir S04846 - UDP-N-acetylmuramoylalanyl-D-glutamyl-L-2,6- diaminopimelate-D-alanyl-D-alanine ligase (EC 6.3.2.15) precursor - Escherichia coli		glycoprotein	1039
694	79313230 (16569, 16570)	Novel Protein sim. G-Bank gil929790 emb CAA88018 - (Z47979) inner membrane or periplasmic protein [Pseudomonas fluorescens]		glycoprotein	1010
695	87898968 (1247, 1248)			glycoprotein	1015
696	32153884 (3221, 3222)			glycoprotein	1044

697	80504143 (4105, 4106)		glycoprotein	1012
698	33206111 (18317, 18318)		glycoprotein	1026
699	28828111 (20577, 20578)		glycoprotein	1008
700	78378895 (9411, 9412)	Novel Protein sim. G-Bank gi 417308 sp P32333 MOT1_YEAST - PROBABLE HELICASE MOT1	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	1003, 1009
701	86694695 (14351, 14352)	Novel Protein sim. G-Bank gi 118411 sp P06634 DED1_YEAST - PUTATIVE ATP-DEPENDENT RNA HELICASE DED1	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1018, 1024
702	79920351 (12087, 12088)	Novel Protein sim. G-Bank gi 132252 sp P24230 RECG_ECOLI - ATP-DEPENDENT DNA HELICASE RECG	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1039
703	20370422 (20703, 20704)	Novel Protein sim. G-Bank gi 1708-95 sp P51979 HFM1_YEAST - HFM1 PROTEIN	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1004
704	20721199 (17441, 17442)	Novel Protein sim. G-Bank gi 2072574 emb CAB083051 - (Z95120) rhlE [Mycobacterium tuberculosis]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1006, 1010
705	20466280 (20809, 20810)	Novel Protein sim. G-Bank gi 2408027 emb CAB16225_1 - (Z99162) putative ATP-dependent RNA helicase	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1010
706	78048670 (10389, 10390)	Novel Protein sim. G-Bank gi 2500537 sp Q03532 HAS1_YEAST - PROBABLE ATP-DEPENDENT RNA HELICASE HAS1	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1008
707	78385461 (7901, 7902)	Novel Protein sim. G-Bank gi 3036880 emb CAA18513 - (AL022374) putative ATP-dependent DNA helicase [Streptomyces coelicolor]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	1006, 1026

708	82457869 (10105, 10106)	Novel Protein sim. GBank gi 3641 emb CAA39465 - (X55993) DBP1 [Saccharomyces cerevisiae]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase@	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1017, 1022, 1023, 1024, 1027, 1034, 1038, 1054
709	78246178 (11645, 11646)	Novel Protein sim. GBank gi 3641 emb CAA39465 - (X55993) DBP1 [Saccharomyces cerevisiae]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1050
710	10103461 (16175, 16176)	Novel Protein sim. GBank gi 3913795 spi Q27830 HELX_METTH - PUTATIVE ATP-DEPENDENT HELICASE MTH1802	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1012
711	13076427 (4097, 4098)	Novel Protein sim. GBank gi 3914617 spi P95122 RECG_MYCTU - ATP- DEPENDENT DNA HELICASE RECG	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase@	1027
712	29255764 (8129, 8130)	Novel Protein sim. GBank gi 421132 pir S30675 - mmtA protein - Escherichia coli	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1022, 1027
713	80245897 (307, 308)	Novel Protein sim. GBank gi 4539568 emb CAB38487.1 - (AL035636) putative helicase [Streptomyces coelicolor]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1006
714	80501629 (18079, 18080)	Novel Protein sim. GBank gi 4539568 emb CAB38487.1 - (AL035636) putative helicase [Streptomyces coelicolor]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1006, 1012, 1017, 1039
715	30268129 (3111, 3112)	Novel Protein sim. GBank gi 4579744 dbj BAA75085.1 - (AB017104) DNA helicase Q1 [Mus musculus]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase@	1026
716	28982349 (12107, 12108)	Novel Protein sim. GBank gi 4579744 dbj BAA75085.1 - (AB017104) DNA helicase Q1 [Mus musculus]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1026
717	30269054 (19419, 19420)	Novel Protein sim. GBank gi 4579744 dbj BAA75085.1 - (AB017104) DNA helicase Q1 [Mus musculus]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1026

718	86693021 (22361, 22362)	Novel Protein sim. G-Bank gj 5102733 emb CAB45191.1 - (AL079292) hypothetical protein, similar to (AC007017) putative RNA helicase A [Arabidopsis thaliana] [Homo sapiens]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1025
719	20297661 (10065, 10066)	Novel Protein sim. G-Bank gj 5669638 gb AAD46404.1 AF096244 - (AF096248) ethylene-responsive RNA helicase [Lycopersicon esculentum]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	helicase	1006, 1022
720	11396799 (1933, 1934)	Novel Protein sim. G-Bank gj 717656 (L00683) - ATP-dependent RNA helicase [Saccharomyces cerevisiae]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1010
721	27823655 (6375, 6376)	Novel Protein sim. G-Bank gj 1723845 sp P53166 YGG4_YEAST - PUTATIVE ATP-DEPENDENT RNA HELICASE YGL064C	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1004
722	27367975 (3559, 3560)	Novel Protein sim. G-Bank gj 1730678 sp P53734 DBP6_YEAST - ATP- DEPENDENT RNA HELICASE DBP6	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1009
723	20466368 (11277, 11278)	Novel Protein sim. G-Bank gj 1731040 sp P54569 YQHH_BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1010
724	80470266 (1591, 1592)	Novel Protein sim. G-Bank gj 2072674 emb CAB08305 - (Z95120) rIE [Mycobacterium tuberculosis]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1003, 1012, 1039
725	16699235 (10817, 10818)	Novel Protein sim. G-Bank gj 2815333 emb CAA16457 - (AL021529) DEAD- box RNA helicase [Streptomyces coelicolor]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1029, 1031, 1039
726	14984445 (13589, 13590)	Novel Protein sim. G-Bank gj 290502 (L10328) - DNA recombinase [Escherichia coli]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1024

727	80063863 (14975, 14978)	Novel Protein sim. GBank gil 5689893 emb [CAB52056.1] - (AL109732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	1022, 1026
728	78768616 (4205, 4206)	Novel Protein sim. GBank gil 3043527 emb [CAA11251] - (AJ223310) helicase [Streptomyces avermitillii]	Contains protein domain (PF00409) - Kinesin light chain repeat	helicase	1017
729	20696804 (17387, 17382)	Novel Protein sim. GBank gil 1070506 pir JBVDH - helicase (EC 3.6.1.-) RADH - yeast [Saccharomyces cerevisiae]		helicase	1034
730	78733980 (21635, 21636)	Novel Protein sim. GBank gil 1070506 pir JBVDH - helicase (EC 3.6.1.-) RADH - yeast [Saccharomyces cerevisiae]		helicase	1026
731	28383392 (17605, 17606)	Novel Protein sim. GBank gil 1074634 pir G64165 - hypothetical protein HI1072 - Haemophilus influenzae (strain Rd KW20)		helicase	1008
732	27926291 (21069, 21070)	Novel Protein sim. GBank gil 1103724 emb [CAA90562] - (Z50182) MgPs [Rhodobacter sphaeroides]		helicase	1022
733	80488958 (9331, 9332)	Novel Protein sim. GBank gil 1169367 sp P45256 DNAB_HAEIN - REPLICATIVE DNA HELICASE		helicase	1012
734	79864561 (3285, 3286)	Novel Protein sim. GBank gil 1169368 sp P46394 DNAB_MYCLE - REPLICATIVE DNA HELICASE HOMOLOG		helicase	1025
735	80025946 (5523, 5524)	Novel Protein sim. GBank gil 1174822 sp Q022322 UVRD_HAEIN - DNA HELICASE II		helicase	1006
736	80222328 (3467, 3468)	Novel Protein sim. GBank gil 172270 (M31524) - PRP16 peptide (put. helicase); putative [Saccharomyces cerevisiae]		helicase	1004, 1006, 1022
737	79417146 (15859, 15860)	Novel Protein sim. GBank gil 1723845 sp P53166 YGG4_YEAST - PUTATIVE ATP-DEPENDENT RNA HELICASE YGL064C		helicase	1006

738 4826)	20202585 (4825, Novel Protein sim. GBank gil1786342 (AE000124) - helicase, ATP-dependent [Escherichia coli]		helicase	1010
739 2382)	35024233 (2381, Novel Protein sim. GBank gil2104219[dbj]BA04879] - (D21853) KIAA0111 [Homo sapiens]		helicase	1022, 1029
740 18506)	80248163 (18505, Novel Protein sim. GBank gil2131434[pir][S70099] hypothetical protein YDR334w - yeast (Saccharomyces cerevisiae)		helicase	1003, 1006, 1008, 1009, 1016, 1017, 1022, 1023, 1024, 1026, 1034, 1039, 1044, 1058
741 5458)	20608443 (5457, Novel Protein sim. GBank gil2245023[emb CAB10443.1] - (Z97341) RNA helicase [Arabidopsis thaliana]		helicase	1004
742 15964)	27824352 (15963, Novel Protein sim. GBank gil2577965[emb CAA75552] - (Y15254) PcfA protein [Bacillus subtilis]		helicase	1004
743 1176)	8756967 (1175, Novel Protein sim. GBank gil2625021 (AF028736) - DNA helicase II [Serratia marcescens]		helicase	1004
744 14472)	20297319 (14471, Novel Protein sim. GBank gil2625021 (AF028736) - DNA helicase II [Serratia marcescens]		helicase	1034
745 16684)	80084142 (16683, Novel Protein sim. GBank gil2916920[emb CAA17667.1] - (AL022004) hypothetical protein Rv0861c [Mycobacterium tuberculosis]		helicase	1022, 1026, 1034
746 7396)	20465199 (7395, Novel Protein sim. GBank gil2984278 (AE000770) - DNA helicase [Aquifex aeolicus]		helicase	1010
747 7780)	20443124 (7799, Novel Protein sim. GBank gil3036880[emb CAA18513] - (AL022374) putative ATP-dependent DNA helicase [Streptomyces coelicolor]		helicase	1024

748	5640527 (973, 974)	Novel Protein sim. GBank gij3047117 (AF056919) - similar to ATP-dependent RNA helicases [Arabidopsis thaliana]		helicase	1058
749	20617248 (15275, 15276)	Novel Protein sim. GBank gij3183525 sp P45018 HRPA_HAEIN - ATP-DEPENDENT HELICASE HRPA HOMOLOG		helicase	1004
750	8468576 (20361, 20362)	Novel Protein sim. GBank gij3183525 sp P45018 HRPA_HAEIN - ATP-DEPENDENT HELICASE HRPA HOMOLOG		helicase	1022
751	11237767 (21089, 21100)	Novel Protein sim. GBank gij3861004 emb CAA14904 - (A J235271) DNA HELICASE II (uvrD) [Rickettsia prowazekii]		helicase	1022
752	10858587 (19863, 19864)	Novel Protein sim. GBank gij3915879 sp Q50629 RUVB_MYCTU - HOLLIDAY JUNCTION DNA HELICASE RUVB		helicase	1031
753	46677315 (21243, 21244)	Novel Protein sim. GBank gij401438 sp P31380 YAB9_YEAST - HYPOTHETICAL 128.5 KD HELICASE IN ATS1-TPD3 INTERGENIC REGION		helicase	1034
754	80192573 (12273, 12274)	Novel Protein sim. GBank gij40500087 (AF109907) - S164 [Homo sapiens]		helicase	1058
755	17893838 (17933, 17934)	Novel Protein sim. GBank gij417308 sp P32333 MOT1_YEAST - PROBABLE HELICASE MOT1		helicase	1006, 1012, 1017, 1022, 1024
756	27361243 (21733, 21734)	Novel Protein sim. GBank gij417308 sp P32333 MOT1_YEAST - PROBABLE HELICASE MOT1		helicase	1009
757	20549482 (5353, 5354)	Novel Protein sim. GBank gij4490937 emb CAB38857.1 - (A J132261) hypothetical helicase K12H4.8-like protein [Homo sapiens]		helicase	1058
758	17670350 (15331, 15332)	Novel Protein sim. GBank gij4490937 emb CAB38857.1 - (A J132261) hypothetical helicase K12H4.8-like protein [Homo sapiens]		helicase	1039

759 6922)	94323888 (6921, Novel Protein sim. GBank gi 4539568 emb CAB38487.1 - (AL035636) putative helicase [Streptomyces coelicolor]		helicase	1003, 1004, 1006, 1008, 1010, 1017, 1022, 1024, 1027, 1031, 1034, 1038, 1054
760 16136)	25145882 (16135, Novel Protein sim. GBank gi 464912 sp P35187 SGS1_YEAST - HELICASE SGS1 (HELICASE TPS1)		helicase	1003
761 3988)	36990620 (3987, Novel Protein sim. GBank gi 5042262 emb CAB44516.1 - (AL078618) putative helicase [Streptomyces coelicolor]		helicase	1016
762 11968)	30670113 (11967, Novel Protein sim. GBank gi 51027733 emb CAB45191.1 - (AL079292) hypothetical protein, similar to (AC007017) putative RNA helicase A [Arabidopsis thaliana] [Homo sapiens]		helicase	1022
763 4654)	56772589 (4653, Novel Protein sim. GBank gi 5678922 gb AAD46808.1 AF15764 - (AF157643) RecB [Mycobacterium smegmatis]		helicase	1021
764 14160)	11697612 (14159, Novel Protein sim. GBank gi 5689893 emb CAB52056.1 - (AL109732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]		helicase	1034
765 17488)	79911905 (17487, Novel Protein sim. GBank gi 585057 sp P37469 DNAC_BACSU - REPLICATIVE DNA HELICASE		helicase	1017, 1044
766 10238)	20297664 (10237, Novel Protein sim. GBank gi 673984 pir HJECDA - helicase (EC 3.6.1.-) IV - Escherichia coli		helicase	1034
767 17512)	20711344 (17511, Novel Protein sim. GBank gi 673985 pir HJNVAV - helicase (EC 3.6.1.-) - Autographa californica nuclear polyhedrosis virus		helicase	1006
768 14032)	80218108 (14031, Novel Protein sim. GBank gi 7311738 sp P38859 DNA2_YEAST - DNA REPLICATION HELICASE DNA2		helicase	1006, 1008, 1022, 1024, 1034

769	80187870 (8475, 8476)	Novel Protein sim. GBank gi 731965 sp P40562 YIS2_YEAST - PUTATIVE ATP-DEPENDENT RNA HELICASE YIR002C		helicase	1058
770	87468970 (10653, 10654)	Novel Protein sim. GBank gi 3142659 (AF064500) - immunoglobulin lambda light chain variable region [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	histone	1015
771	80236509 (17025, 17026)	Novel Protein sim. GBank gi 1219761 sp P04912 H2A2_YEAST - HISTONE H2A.2	Contains protein domain (PF00125) - Core histone H2A/H2B/H3/H4	histone	1006, 1008, 1010, 1017, 1022, 1023, 1025, 1038, 1039
772	79242158 (9259, 9260)	Novel Protein sim. GBank gi 729671 sp P40280 H2A_MAIZE - HISTONE H2A	Contains protein domain (PF00125) - Core histone H2A/H2B/H3/H4	histone	1003, 1017, 1044
773	30577588 (9807, 9808)	Novel Protein sim. GBank gi 731418 sp P39984 HAT2_YEAST - HISTONE ACETYLTRANSFERASE TYPE B SUBUNIT 2	Contains protein domain (PF00400) - WD domain, G- beta repeat	histone	1006, 1026
774	80221135 (2679, 2680)	Novel Protein sim. GBank gi 123373 sp P26584 HMG2_CHICK - HIGH MOBILITY GROUP PROTEIN HMG2 (HMG-2)	Contains protein domain (PF00505) - HMG (high mobility group) box	histone	1006, 1022, 1034
775	80421194 (2935, 2936)	Novel Protein sim. GBank gi 417699 sp P32561 RPD3_YEAST - HISTONE DEACETYLYLASE RPD3 (TRANSCRIPTIONAL REGULATORY PROTEIN RPD3)	Contains protein domain (PF00850) - Histone deacetylase family	histone	1003, 1006, 1022, 1024, 1025, 1038, 1044
776	78971427 (8901, 8902)	Novel Protein sim. GBank gi 1708096 sp P53551 H1_YEAST - HISTONE H1		histone	1039
777	25176959 (3713, 3714)	Novel Protein sim. GBank gi 2257756 (U82815) - nucleolar histone deacetylase HD2-p39 [Zea mays]		histone	1003
778	28396269 (12785, 12786)	Novel Protein sim. GBank gi 249843 sp Q12341 HAT1_YEAST - HISTONE ACETYLTRANSFERASE		histone	1006, 1038
779	94666912 (8277, 8278)	Novel Protein sim. GBank gi 417038 sp Q03330 GCN5_YEAST - TRANSCRIPTIONAL ACTIVATOR GCN5		histone	1024

780	5706086 (13465, 13466)	Novel Protein sim. GBank gi 417689 sp P32561 RPD3_YEAST - HISTONE DEACETYLASE RPD3 (TRANSCRIPTIONAL REGULATORY PROTEIN RPD3)		histone	1058
781	65495666 (22847, 22848)	Novel Protein sim. GBank gi 417699 sp P32561 RPD3_YEAST - HISTONE DEACETYLASE RPD3 (TRANSCRIPTIONAL REGULATORY PROTEIN RPD3)		histone	1023
782	78727051 (6121, 6122)	Novel Protein sim. GBank gi 462231 sp P35062 H2A3_CHICK - HISTONE H2A-III		histone	1008
783	80219184 (19783, 19784)	Novel Protein sim. GBank gi 731418 sp P39884 HAT2_YEAST - HISTONE ACETYLTRANSFERASE_TYPE B SUBUNIT 2		histone	1006, 1022
784	13042444 (8621, 8622)	Novel Protein sim. GBank gi 526251 emb CAB45706.1 - (AL080091) hypothetical protein [Homo sapiens]	Contains protein domain (PF00023) - Ank repeat	homeobox	1024
785	65462295 (887, 888)	Novel Protein sim. GBank gi 1708053 sp P54366 GSC_DROME - HOMEobox PROTEIN GOOSECOID	Contains protein domain (PF00046) - Homeobox domain	homeobox	1054
786	57535052 (11635, 11636)	Novel Protein sim. GBank gi 731599 sp P38744 YHB8_YEAST - HYPOTHETICAL 14.0 KD PROTEIN IN APM2-DUR3 INTERGENIC REGION	Contains protein domain (PF01329) - Pterin 4 alpha carbolinolamine dehydratase	homeobox	1003, 1008, 1009, 1022, 1034, 1044, 1058
787	17941386 (18161, 18162)	Novel Protein sim. GBank gi 1346387 sp P48000 KNA3_ARATH - KNOTTED LIKE HOMEBOX PROTEIN 3		homeobox	1008
788	80079864 (22349, 22350)	Novel Protein sim. GBank gi 285481 sp P37631 YHN_ECOLI - HYPOTHETICAL 43.8 KD PROTEIN IN RHSB-PIT INTERGENIC REGION		homeobox	1016, 1022, 1034
789	16841471 (22513, 22514)	Novel Protein sim. GBank gi 2985575 gb AAC32818 - (AF050181) KNOX class homeodomain protein [Oryza sativa]		homeobox	1044

790	27982094 (8479, 8480)		homeobox	1053
791	10355887 (7623, 7624)	Novel Protein sim. GBank gil3695013 (AF052586) - CtrA [Pseudomonas aeruginosa]	Contains protein domain (PF00142) - 4Fe-4S iron sulfur cluster binding proteins. NifH/IfrxC family	1001
792	80572341 (471, 472)	Novel Protein sim. GBank gil1206013 (U44087) - beta-D-glucosidase precursor [Zea mays]	Contains protein domain (PF00232) - Glycosyl hydrolase family 1	1003, 1058
793	80063054 (11951, 11952)	Novel Protein sim. GBank gil2642340 (AF032970) - imidazolone propionate hydrolase [Pseudomonas putida]	Contains protein domain (PF00449) - Urease	1024
794	94993457 (17843, 17844)	Novel Protein sim. GBank gil586023 sp P37819 SPPEB_STRL - POSSIBLE AGMATINASE (AGMATINE UREOHYDROLASE) (AUH) (PROCLAVAMINIC ACID AMIDINO HYDROLASE)	Contains protein domain (PF00491) - Arginase family	1004, 1006, 1010, 1022, 1024, 1034
795	20451078 (1973, 1974)	Novel Protein sim. GBank gil728887 sp P40906 ARGI_COClM - ARGINASE	Contains protein domain (PF00491) - Arginase family	1024
796	17384473 (12687, 12688)	Novel Protein sim. GBank gil2499911 sp Q64230 MEPA_RAT - MEPRIN A ALPHA-SUBUNIT PRECURSOR (ENDOPEPTIDASE-2) (MEP-1) (ENDOPEPTIDASE-24.18 ALPHA-SUBUNIT) (E-24.18)	Contains protein domain (PF00917) - MATH domain	1003
797	30654736 (13343, 13344)	Novel Protein sim. GBank gil280870 emb CAA16228 - (AL021426) cwlM [Mycobacterium tuberculosis]	Contains protein domain (PF01471) - Putative peptidoglycan binding domain	1022
798	54994464 (11235, 11236)	Novel Protein sim. GBank gil4106521 gb AAD02873.11 - (AF097909) fibril-like structure subunit FibA [Peptostreptococcus micros]	Contains protein domain (PF01473) - Putative cell wall binding domain	1029

799	78378458 (11907, 11908)	Novel Protein sim. GBank gi 5596789 emb CAB51436.1 - (AL096884) putative phosphoribosyl-AMP cyclohydrolase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01502) - Phosphoribosyl- AMP cyclohydrolase	hydrolase	1026, 1029
800	80079417 (21419, 21420)	Novel Protein sim. GBank gi 7297171 sp P06989 HIS2_ECOLI - PHOSPHORIBOSYL-AMP CYCLOCYDROLASE / PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE	Contains protein domain (PF01502) - Phosphoribosyl- AMP cyclohydrolase	hydrolase	1034
801	20396135 (3553, 3554)	Novel Protein sim. GBank gi 3026224 sp P71505 DLHH_METEX - PUTATIVE CARBOXYMETHYLENEBUTENOLIDASE (DIENELACTONE HYDROLASE) (DLH)	Contains protein domain (PF01738) - Dienelactone hydrolase family	hydrolase	1004
802	95354023 (9013, 9014)	Novel Protein sim. GBank gi 3913498 sp Q43914 DLHH_AZOBR - PUTATIVE CARBOXYMETHYLENEBUTENOLIDASE (DIENELACTONE HYDROLASE) (DLH)	Contains protein domain (PF01738) - Dienelactone hydrolase family	hydrolase	1058
803	80503570 (13161, 13162)	Novel Protein sim. GBank gi 1072325 emb CAA56560 - (X80276) AmpS beta-factamase [Aeromonas sobria]		hydrolase	1012
804	20467785 (11159, 11160)	Novel Protein sim. GBank gi 1076211 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii		hydrolase	1010
805	37008433 (4767, 4768)	Novel Protein sim. GBank gi 1109800 dbj BAA04049 - (D16628) ATSeH [Arabidopsis thaliana]		hydrolase	1029
806	29449184 (7813, 7814)	Novel Protein sim. GBank gi 113310 sp P15937 ACU8_NEUCR - ACETYL- COA HYDROLASE (ACETYL-COA DEACYLASE) (ACETYL-COA ACYLASE) (ACETATE UTILIZATION PROTEIN)		hydrolase	1026

807	29345082 (19039, 19040)	Novel Protein sim. GBank gi 113310 sp P15937 ACU8_NEUCR - ACETYL-COA HYDROLASE (ACETYL-COA ACYLASE) (ACETATE UTILIZATION PROTEIN)	hydrolase	1009, 1026
808	20720739 (12927, 12928)	Novel Protein sim. GBank gi 1168388 sp P43473 AGL_PEDPE - ALPHA-GLUCOSIDASE (MALTASE)	hydrolase	1006
809	66373053 (12709, 12710)	Novel Protein sim. GBank gi 1168446 sp P45274 AMPN_HAEIN - AMINOPEPTIDASE N (ALPHA-AMINOACYLPEPTIDE HYDROLASE)	hydrolase	1036
810	20430082 (747, 748)	Novel Protein sim. GBank gi 1169532 sp P43806 ENO_HAEIN - ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)	hydrolase	1010
811	86665299 (19423, 19424)	Novel Protein sim. GBank gi 1170427 sp P25503 HUTU_BACSU - UROCANATE HYDRATASE (UROCANASE) (IMIDAZOLENEPROPIONATE HYDROLASE)	hydrolase	1019
812	30385959 (17775, 17776)	Novel Protein sim. GBank gi 124867 sp P28239 IPY2_YEAST - INORGANIC PYROPHOSPHATASE, MITOCHONDRIAL PRECURSOR (PYROPHOSPHATE PHOSPHOHYDROLASE) (PPASE)	hydrolase	1022
813	85801171 (14143, 14144)	Novel Protein sim. GBank gi 124870 sp P17288 PYR_ECOLI - INORGANIC PYROPHOSPHATASE (PYROPHOSPHATE PHOSPHOHYDROLASE) (PPASE)	hydrolase	1025
814	13527697 (21599, 21600)	Novel Protein sim. GBank gi 129001 sp P13398 NYLA_PSES8 - 6-AMINOHEXANOATE-CYCLIC-DIMER HYDROLASE (NYLON OLIGOMERS DEGRADING ENZYME EI)	hydrolase	1001

815 199676)	27980115 (19675, Novel Protein sim. GBank gi 1582641 p0f 2119210A - mucin [Homo sapiens]		hydrolase	10066
816 21166)	Novel Protein sim. GBank gi 1653547 dbj BAA18460 - (D90914) hypothetical protein [Synechocystis sp.]		hydrolase	1008, 1010, 1017, 1044
817 9608)	Novel Protein sim. GBank gi 1703266 sp Q11056 AM12_MYCTU - PUTATIVE AMIDASE CY50_19C		hydrolase	1003, 1039
818 16060)	Novel Protein sim. GBank gi 1730265 sp P49851 YKHA_BACSU - HYPOTHETICAL 20.1 KD PROTEIN IN HMP 5'REGION (ORF1)		hydrolase	1017
819 9000)	Novel Protein sim. GBank gi 1731316 sp P54955 YXEP_BACSU - HYPOTHETICAL 41.6 KD PROTEIN IN IDH- DEOR INTERGENIC REGION		hydrolase	1012
820 17764)	Novel Protein sim. GBank gi 1871461 dbj BAA12150 - (D83955) 2-hydroxy-6- oxo-7-methylocta-2,4-dienoate hydrolase [Pseudomonas fluorescens]		hydrolase	1001
821 18400)	Novel Protein sim. GBank gi 2494852 sp Q12320 GLO4 YEAST - HYDROXYACYLGLUTATHIONE HYDROLASE ISOZYME 2 (GLYOXALASE II) (GLX II)		hydrolase	1022
822 84668)	Novel Protein sim. GBank gi 2494855 sp P72933 GLO2_SYNY3 - PROBABLE HYDROXYACYLGLUTATHIONE HYDROLASE (GLYOXALASE II) (GLX II)		hydrolase	1024
823 8582)	Novel Protein sim. GBank gi 2499987 sp P96386 PTH_MYCTU - PEPTIDYL- TRNA HYDROLASE (PTH)		hydrolase	1004
824 21822)	Novel Protein sim. GBank gi 2791507 emb CAA16044 - (AL021246) pepD [Mycobacterium tuberculosis]		hydrolase	1029

825	11766387 (353, 354)	Novel Protein sim. GBank gil 2880042 (AC002340) - putative 3-hydroxyisobutyryl-coenzyme A hydrolase [Arabidopsis thaliana]		hydrolase	1016
826	79220388 (823, 824)	Novel Protein sim. GBank gil 3023706 sim P96377 ENO_MYCTU - ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)		hydrolase	1004, 1008, 1029
827	78784708 (17385, 17386)	Novel Protein sim. GBank gil 3860903 sim CAA14803 - (AJ235271) unknown [Rickettsia prowazekii]		hydrolase	1009
828	16841055 (16927, 16928)	Novel Protein sim. GBank gil 409365 (M33923) - urocanase [Pseudomonas putida]		hydrolase	1023
829	78526490 (20447, 20448)	Novel Protein sim. GBank gil 409365 (M33923) - urocanase [Pseudomonas putida]		hydrolase	1008, 1026, 1036
830	94134179 (11807, 11808)	Novel Protein sim. GBank gil 4107260 sim CAA09772 - (AJ011781) amylosucrase [Neisseria polysaccharea]		hydrolase	1010
831	80497359 (13003, 13004)	Novel Protein sim. GBank gil 467250 sim CAB37575 - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]		hydrolase	1006, 1010, 1012, 1016, 1022, 1034
832	34082758 (16417, 16418)	Novel Protein sim. GBank gil 4583153 gb AAD24964_1 - (AF108211) cytosolic inorganic pyrophosphatase [Homo sapiens]		hydrolase	1000, 1029, 1034
833	13527799 (6275, 6276)	Novel Protein sim. GBank gil 4887205 gb AAD32232_1 AF14744 - (AF147448) membrane-bound lytic transglycosylase [Pseudomonas aeruginosa]		hydrolase	1027, 1039
834	47654633 (13637, 13638)	Novel Protein sim. GBank gil 4980556 gb AAD35163_1 AE00169 - (AE001693) D-mannonate hydrolase [Thermotoga maritima]		hydrolase	1029
835	66582086 (14257, 14258)	Novel Protein sim. GBank gil 5091484 gb AAD39557_1 - (AF031417) PcaC-like protein [Pseudomonas putida]		hydrolase	1009, 1023

836	94140415 (18437, 18438)	Novel Protein sim. GBank g 513962 gb CAE45554.1 - (AL079353) putative epoxide hydrolase [Streptomyces coelicolor]		hydrolase	1003
837	39484694 (3061, 3062)	Novel Protein sim. GBank g 5354194 gb AAD42403.1 AF15749 - (AF157493) membrane stanlyl aminopeptidase [Zymomonas mobilis]		hydrolase	1004
838	80427935 (11453, 11454)	Novel Protein sim. GBank g 584749 sp P37112 AMAA_BACST - N-ACYL-L-AMINO ACID AMIDOHYDROLASE (AMINOACYLASE)		hydrolase	1025
839	24111183 (14367, 14368)	Novel Protein sim. GBank g 731088 sp P24215 UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		hydrolase	1044
840	29680154 (22185, 22186)	Novel Protein sim. GBank g 731088 sp P24215 UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		hydrolase	1026
841	79250832 (21213, 21214)	Novel Protein sim. GBank g 731877 sp P40462 YIN7_YEAST - HYPOTHETICAL ZINC AMINOPEPTIDASE YIL137C		hydrolase	1003, 1008, 1023, 1039
842	11351308 (18389, 18390)	Novel Protein sim. GBank g 568397 U06274 - UDP-glucuronosyltransferase [Rattus norvegicus]	Contains protein domain (PF00201) - UDP-glucuronosyl and UDP-glucosyl transferases	hydroxysteroid	1034
843	28800645 (5701, 5702)	Novel Protein sim. GBank g 135033 sp P07631 SUH2_RAT - PROBABLE ALCOHOL SULFOTRANSFERASE (HYDROXYSTEROID SULFOTRANSFERASE) (ST) (SENESCENCE MARKER PROTEIN 2) (SMP-2) (ANDROGEN-REPRESSIBLE LIVER PROTEIN) (DEHYDROEPIANDROSTERONE SULFOTRANSFERASE) (DST)		hydroxysteroid	1039

844	79563276 (6491, 6492)	Novel Protein sim. GBank gj 5030426 gb AAC16243.2 - (AF061748) cell division protein FtsH [Streptococcus pneumoniae]		hypoxanthine	1026, 1039
845	86677700 (21015, 21016)	Novel Protein sim. GBank gj 106664 pir S19663 - Ig lambda chain V region (clone alpha-BSA3) - human	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1048
846	6430994 (16511, 16512)	Novel Protein sim. GBank gj 185085 (L12078) - immunoglobulin heavy chain [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1058
847	3246449 (13303, 13304)	Novel Protein sim. GBank gj 1871492 dbj BAA12058 - (D83691) IgM [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1025
848	55777005 (3183, 3184)	Novel Protein sim. GBank gj 2135457 pir 37780 - Ig variable region (V/D) (clone T20-11) - human (fragment)	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1021
849	56066044 (16843, 16844)	Novel Protein sim. GBank gj 2345012 (U96283) - Ig heavy chain VH4 region [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1021
850	21423498 (2729, 2730)	Novel Protein sim. GBank gj 284186 pir S26909 - Ig heavy chain V region (DP-3) - human (fragment)	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1022
851	88098275 (2443, 2444)	Novel Protein sim. GBank gj 292199 (L12187) - immunoglobulin mu chain [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1015
852	87449755 (69, 70)	Novel Protein sim. GBank gj 3091184 (AF060135) - immunoglobulin lambda light chain variable region [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1013
853	55776576 (14215, 14216)	Novel Protein sim. GBank gj 33478 emb CAA49749 - (X70208) immunoglobulin M chain [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1021
854	35994043 (12629, 12630)	Novel Protein sim. GBank gj 4107050 dbj BAA36310 - (AB021518) immunoglobulin heavy chain variable region (IgM) [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1024

855	55309289 (11557, 11558)	Novel Protein sim. GBank gil420059 pir A45953 - Ig heavy chain precursor V-III region (VH26) - human	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1019
856	20479843 (6579, 6580)	Novel Protein sim. GBank gil789251 (U24689) - Immunoglobulin heavy chain variable region [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	immunoglob	1024
857	11221443 (3453, 3454)	Novel Protein sim. GBank gil1791181 (U80169) - Immunoglobulin heavy chain variable region [Homo sapiens]		immunoglob	1022
858	55776583 (15927, 15928)	Novel Protein sim. GBank gil186233 (M83134) - Immunoglobulin heavy chain V-region [Homo sapiens]		immunoglob	1021
859	77757436 (17471, 17472)	Novel Protein sim. GBank gil229073 pir 1817343A - Ig M variable region:SUBUNIT=heavy chain [Homo sapiens]		immunoglob	1021
860	79323799 (11127, 11128)	Novel Protein sim. GBank gil2344984 (U83195) - IgM heavy chain VH1 region precursor [Homo sapiens]		immunoglob	1022
861	91225598 (3191, 3192)	Novel Protein sim. GBank gil284189 pir S26928 - Ig heavy chain V region (DP-32) - human (fragment)		immunoglob	1018
862	85818282 (10961, 10962)	Novel Protein sim. GBank gil3170705 (AF062121) - Immunoglobulin heavy chain variable region [Homo sapiens]		immunoglob	1019
863	800212582 (19325, 19326)	Novel Protein sim. GBank gil3170765 (AF062151) - Immunoglobulin heavy chain variable region [Homo sapiens]		immunoglob	1024
864	800062616 (11617, 11618)	Novel Protein sim. GBank gil3171071 (AF062299) - Immunoglobulin heavy chain variable region [Homo sapiens]		immunoglob	1010
865	555504106 (20231, 20232)	Novel Protein sim. GBank gil346082 pir PH1247 - Ig heavy chain V region (clone CORD5) - human (fragment)		immunoglob	1021
866	95010182 (21717, 21718)	Novel Protein sim. GBank gil5305639 gb AAD1751.1 - (AF109215) TibA [Escherichia coli]		immunoglob	1039

867 14996)	86381147 (14965, Novel Protein sim. GBank gil90321 pir PS0401 - basigin type II - mouse (fragment)		immunoglob	1053
868 15590)	80217714 (15889, Novel Protein sim. GBank gil1706485 sp P5486 DNM1 _YEAST - DYNAMIN- RELATED PROTEIN DNM1	Contains protein domain (PF00350) - Dynamin family	interferon	1003, 1004, 1006, 1009, 1022, 1023, 1024, 1026, 1031, 1034, 1038, 1039
869 38444220 (21351, 21352)	38444220 (21351, Novel Protein sim. GBank gil111876 pir JC1241 - beta-interferon-induced protein - rat		interferon	1034, 1044
870 22256)	11290053 (22255, Novel Protein sim. GBank gil400060 sp Q01628 INIU_HUMAN - INTERFERON-INDUCIBLE PROTEIN 1-8U		Interferon	1004
871 9806)	35061470 (9805, Novel Protein sim. GBank gil570321 0 emb CAB52384.1 - (Y07829) RING finger protein [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	interleukinrecept	1038
872 4970)	27372344 (4969, Novel Protein sim. GBank gil1403700 (U48592) - interleukin-1 receptor accessory protein [Rattus norvegicus]		interleukinrecept	1009
873 11966)	94126308 (11965, Novel Protein sim. GBank gil4504655 ref NP_002177.1 pir 9R - interleukin 9 receptor		interleukinrecept	1030, 1053
874 16886)	79492599 (16885, Novel Protein sim. GBank gil1532175 (U63815) - similar to protein disulfide isomerase [Arabidopsis thaliana]	Contains protein domain (PF00085) - Thioredoxin	isomerase	1004
875 2080)	79239223 (2079, Novel Protein sim. GBank gil2501208 sp Q14554 PDIR_HUMAN - PROTEIN DISULFIDE ISOMERASE-RELATED PROTEIN PRECURSOR (PDIR)	Contains protein domain (PF00085) - Thioredoxin	isomerase	1026, 1044
876 1776)	38925255 (1775, Novel Protein sim. GBank gil136063 sp P12863 TPIS_MAIZE - TRIOSEPHOSPHATE ISOMERASE, CYTOSOLIC (Tim)	Contains protein domain (PF00121) - Triosephosphate isomerase	isomerase	1008

877	13086040 (6799, 6800)	Novel Protein sim. GBank gil2671541 sp P19583 TPIS_CORGL - TRIOSEPHOSPHATE ISOMERASE (TIM)	Contains protein domain (PF00121) - Triosephosphate isomerase	isomerase	1027
878	66002656 (10469, 10470)	Novel Protein sim. GBank gil11808 sp P24367 CYPB_CHICK - PEPTIDYL-PROLYL CIS-TRANS ISOMERASE B PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN B) (S-CYCLOPHILIN) (SCYLP)	Contains protein domain (PF00160) - Cyclophilin type peptidyl-prolyl cis-trans isomerase	isomerase	1006, 1010
879	28389536 (21877, 21878)	Novel Protein sim. GBank gil1181041 sp P21569 CYPH_MAIZE - PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (PPIASE) (ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN)	Contains protein domain (PF00160) - Cyclophilin type peptidyl-prolyl cis-trans isomerase	isomerase	1044
880	79582941 (7407, 7408)	Novel Protein sim. GBank gil1181141 sp P25334 CYPY_YEAST - PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE) (ROTAMASE)	Contains protein domain (PF00160) - Cyclophilin type peptidyl-prolyl cis-trans isomerase	isomerase	1027
881	39573486 (6911, 6912)	Novel Protein sim. GBank gil4619011 sp P35176 CYPD_YEAST - PEPTIDYL-PROLYL CIS-TRANS ISOMERASE D PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN D)	Contains protein domain (PF00160) - Cyclophilin type peptidyl-prolyl cis-trans isomerase	isomerase	1022
882	13055250 (3259, 3260)	Novel Protein sim. GBank gil153085 (M86227) - DNA gyrase B subunit [Staphylococcus aureus]	Contains protein domain (PF00204) - DNA topoisomerase II (N-terminal region)	isomerase	1024
883	11084281 (19719, 19720)	Novel Protein sim. GBank gil629198 pir S44198 - DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - Mycobacterium tuberculosis	Contains protein domain (PF00204) - DNA topoisomerase II (N-terminal region)	isomerase	1024
884	11100397 (20569, 20570)	Novel Protein sim. GBank gil629198 pir S44198 - DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - Mycobacterium tuberculosis	Contains protein domain (PF00204) - DNA topoisomerase II (N-terminal region)	isomerase	1006, 1023

885	88095012 (9299, 9300)	Novel Protein sim. GBank gil120226 sp P28725 FKBP STRCH - FK506-BINDING PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PPIASE) (ROTAMASE)	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	1010, 1012, 1022, 1024, 1034
886	21144943 (21875, 21876)	Novel Protein sim. GBank gil2104959 (U96925) - immunophilin [Vicia faba]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	1010, 1034, 1044
887	79420957 (12533, 12534)	Novel Protein sim. GBank gil633644 emb CAA84280 - (234523) FKBP-33 [Streptomyces chrysomallus]	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	1006, 1016
888	79185742 (18659, 18660)	Novel Protein sim. GBank gil1175033 sp P44398 XYLA_HAEIN - XYLOSE ISOMERASE	Contains protein domain (PF00259) - Xylose isomerase	isomerase	1022, 1027
889	27965881 (91119, 9120)	Novel Protein sim. GBank gil234318 bbs S57887 - (S57886) Isopropylmalate isomerase {EC 4.2.1.33} [Saccharomyces cerevisiae, Peptide, 782_aal] [Saccharomyces cerevisiae]	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	isomerase	1034
890	17897355 (16753, 16754)	Novel Protein sim. GBank gil2492639 sp Q44427 LEU2_ACTT1 - 3-ISOPROPYLMALATE DEHYDRATASE LARGE SUBUNIT (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI)	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	isomerase	1024
891	32160031 (13381, 13382)	Novel Protein sim. GBank gil482062 pif S40586 - 3-isopropylmalate dehydratase (EC 4.2.1.33) alpha chain - Escherichia coli	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	isomerase	1023
892	27843543 (22931, 22932)	Novel Protein sim. GBank gil341351 emb CAA03982 - (AJ000265) glucose-6-phosphate isomerase [Spinacia oleracea]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	1022

893	8487783 (6949, 6950)	Novel Protein sim. G-Bank gi 4468678 emb CAB38132.1 - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	1010
894	80021176 (12405, 12406)	Novel Protein sim. G-Bank gi 4468678 emb CAB38132.1 - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	1006, 1008, 1010, 1029, 1054
895	78914208 (18703, 18704)	Novel Protein sim. G-Bank gi 1173139 sp P46989 RPE_YEAST - RIBULOSE-PHOSPHATE 3-EPIMERASE (FENTOSE-5-PHOSPHATE 3-EPIMERASE) (RPE) (RPE)	Contains protein domain (PF00834) - Ribulose-phosphate 3 epimerase family	isomerase	1039
896	80056947 (9951, 9952)	Novel Protein sim. G-Bank gi 629198 p S44198 - DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - Mycobacterium tuberculosis	Contains protein domain (PF00986) - DNA gyrase B subunit, carboxyl terminus	isomerase	1004, 1022, 1023
897	80026840 (13431, 13432)	Novel Protein sim. G-Bank gi 2352095 (U97022) - DNA topoisomerase I [Fervidobacterium islandicum]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	1024
898	144011555 (7763, 7764)	Novel Protein sim. G-Bank gi 415510 (AE001522) - topoisomerase I [Helicobacter pylori J99]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	1053
899	70959417 (11553, 11554)	Novel Protein sim. G-Bank gi 4539561 emb CAB38480.1 - (AL035636) probable DNA topoisomerase I [Streptomyces coelicolor]	Contains protein domain (PF01131) - Prokaryotic DNA topoisomerase	isomerase	1029
900	82486742 (15805, 15806)	Novel Protein sim. G-Bank gi 729984 sp P29952 MANA_YEAST - MANNOSE-6-PHOSPHATE ISOMERASE (PHOSPHOMANNOSE ISOMERASE) (PMI) (PHOSPHOHEXOMUTASE)	Contains protein domain (PF01238) - Phosphomannose isomerase type I	isomerase	1006, 1010, 1022, 1024, 1026, 1029, 1034
901	11139967 (12247, 12248)	Novel Protein sim. G-Bank gi 1248381 sp P15496 IPPL_YEAST - ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (IPP ISOMERASE)	Contains protein domain (PF01772) - Isopentenyl-diphosphate delta-isomerase	isomerase	1017, 1022, 1024

902	80079053 (18617, 18618)	Novel Protein sim. GBank gi 116841 sp P21640 COBJ_PSEDE - PRECORRIN-3B C17-METHYLTRANSFERASE (PRECORRIN-3 METHYLTRANSFERASE) (PRECORRIN-3 METHYLASE)		isomerase	1034
903	11398800 (21731, 21732)	Novel Protein sim. GBank gi 1173141 sp P44725 RPIA_HAEIN - RIBOSE 5-PHOSPHATE ISOMERASE A (PHOSPHORIBOSOMERASE A)		isomerase	1034
904	78735341 (5697, 5698)	Novel Protein sim. GBank gi 119120 sp P07896 ECHP_RAT - ENOYL-COA HYDRATASE / 3,2-TRANS-ENOYL-COA ISOMERASE / 3-HYDROXYACYL-COA DEHYDROGENASE, PEROXISOMAL (PEROXISOMAL BIFUNCTIONAL ENZYME) (PBE) (PBFE)		isomerase	1022
905	36989019 (4049, 4050)	Novel Protein sim. GBank gi 1346229 sp P48369 GYRA_AERSA -DNA GYrase subunit A		isomerase	1016
906	87114770 (2447, 2448)	Novel Protein sim. GBank gi 1346235 sp Q07702 GYRA_MYCTU - DNA GYrase subunit A		isomerase	1038
907	11696770 (20035, 20036)	Novel Protein sim. GBank gi 1346235 sp Q07702 GYRA_MYCTU - DNA GYrase subunit A		isomerase	1006
908	79859058 (6961, 6962)	Novel Protein sim. GBank gi 1346237 sp P48372 GYRA_PSEAE - DNA GYrase subunit A		isomerase	1023
909	80029852 (115713, 15714)	Novel Protein sim. GBank gi 1346237 sp P48372 GYRA_PSEAE - DNA GYrase subunit A		isomerase	1022, 1034
910	47660437 (5665, 5666)	Novel Protein sim. GBank gi 1350839 sp P47968 RPIA_MOUSE - RIBOSE 5-PHOSPHATE ISOMERASE (PHOSPHORIBOSOMERASE)		isomerase	1029

911	80219119 (7069, 7070)	Novel Protein sim. GBank gil1493831 (U47282) - Trf5p [Saccharomyces cerevisiae]		isomerase	1003, 1004, 1006, 1008, 1017, 1022, 1024, 1034, 1039
912	27827914 (13299, 13300)	Novel Protein sim. GBank gil1718068 sp P42607 UXAC_ECOLI - URONATE ISOMERASE (GLUCURONATE ISOMERASE) (URONIC ISOMERASE)		isomerase	1004
913	78795281 (20327, 20328)	Novel Protein sim. GBank gil173004 (K03077) - topoisomerase I [Saccharomyces cerevisiae]		isomerase	1026
914	52563108 (19509, 19510)	Novel Protein sim. GBank gil1730672 sp P53728 YN8G_YEAST - PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE YNR028W		isomerase	1034
915	17941397 (20003, 20004)	Novel Protein sim. GBank gil1786248 gb AAC73173.1 - (AE000116) L-arabinose isomerase [Escherichia coli]		isomerase	1039
916	79756270 (6919, 6920)	Novel Protein sim. GBank gil2072722 emb CA08326 - (Z95121) manA [Mycobacterium tuberculosis]		isomerase	1004
917	13005366 (13157, 13158)	Novel Protein sim. GBank gil2072722 emb CA08326 - (Z95121) manA [Mycobacterium tuberculosis]		isomerase	1027
918	78788082 (9643, 9644)	Novel Protein sim. GBank gil2144771 p JC4557 - capreomycin acetyltransferase (EC 2.3.1.) - Streptomyces capreolus		isomerase	1009
919	79481083 (277, 278)	Novel Protein sim. GBank gil2494591 sp P80449 FOLX_ECOLI - D-ERYTHRO-7-8-DIHYDROOPTERIN TRIPHOSPHATE EPIMERASE (DIHYDROOPTERIN TRIPHOSPHATE 2'-EPIMERASE)		isomerase	1022

920	19900373 (18937, 18938)	Novel Protein sim. GBank g 2494660 sp Q45291 GALE_BRELA - UDP- GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4- EPIMERASE)		isomerase	1024
921	78471498 (10871, 10872)	Novel Protein sim. GBank g 2495634 sp P76516 YFDE_ECOLI - HYPOTHETICAL 43.3 KD PROTEIN IN EVGS- GLK INTERGENIC REGION		isomerase	1026
922	80254034 (9247, 9248)	Novel Protein sim. GBank g 2498994 sp Q54177 TGT_SH-FL - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA- GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME) (VIRULENCE- ASSOCIATED PROTEIN VACC)		isomerase	1034
923	80504157 (6963, 6964)	Novel Protein sim. GBank g 2764829 emb CAA66096 - (X97452) enoyl-CoA Isomerase [Escherichia coli]		isomerase	1012
924	47660019 (15855, 15856)	Novel Protein sim. GBank g 2791505 emb CAA16042 - (AL021246) rpi [Mycobacterium tuberculosis]		isomerase	1029
925	95292984 (1015, 1016)	Novel Protein sim. GBank g 2983605 (AE000725) - ribose 5-phosphate isomerase B [Aquifex aeolicus]		isomerase	1001, 1004, 1010, 1025, 1027, 1029, 1054
926	20726453 (21161, 21162)	Novel Protein sim. GBank g 3393015 emb CAA20107 - (AL031174) dna topoisomerase ii [Schizosaccharomyces pombe]		isomerase	1006
927	9681086 (17099, 17100)	Novel Protein sim. GBank g 3820518 (AF061751) - isomerase PhnD [Burkholderia sp. RP007]		isomerase	1039
928	19890424 (10835, 10836)	Novel Protein sim. GBank g 3914532 sp O32054 QUEA_BACSU - S- ADENOSYLMETHIONINE:TRNA RIBOSYLTRANSFERASE-ISOMERASE (QUEUOSINE BIOSYNTHESIS PROTEIN QUEA)		isomerase	1054

929	66378371 (16897, 16898)	Novel Protein sim. GBank gi 413835 emb CAA06715 - (AJ005815) DNA gyrase subunit A [Streptococcus pneumoniae]		isomerase	1008
930	79863174 (4427, 4428)	Novel Protein sim. GBank gi 4176379 gb BAA37152 - (AB003428) topoisomerase IV subunit [Pseudomonas aeruginosa]		isomerase	1017
931	70336480 (11873, 11874)	Novel Protein sim. GBank gi 4582393 emb CAB40340.1 - (AL049573) putative precorin-8X methyltransferase [Streptomyces coelicolor]		isomerase	1049
932	20465331 (1219, 1220)	Novel Protein sim. GBank gi 543671 sp P35673 GALE_ERWAM - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	1010
933	80257881 (12407, 12408)	Novel Protein sim. GBank gi 544465 sp P35885 GYRA_STRCO - DNA GYRASE SUBUNIT A		isomerase	1003, 1022
934	20632843 (12129, 12130)	Novel Protein sim. GBank gi 5459388 emb CAB50746.1 - (AL096639) putative aminotransferase [Streptomyces coelicolor]		isomerase	1004
935	65660819 (4901, 4902)	Novel Protein sim. GBank gi 5616307 gb AAD45716.1 - (AF160811) L-ribulose 5-phosphate 4-epimerase [Bacillus stearothermophilus]		isomerase	1041
936	27981232 (5189, 5190)	Novel Protein sim. GBank gi 732387 sp P39156 YWLF_BACSU - HYPOTHETICAL LACA/RPIB FAMILY PROTEIN IN SPOIIR-GLYC INTERGENIC REGION		isomerase	1006
937	10054014 (20831, 20832)	Novel Protein sim. GBank gi 732387 sp P39156 YWLF_BACSU - HYPOTHETICAL LACA/RPIB FAMILY PROTEIN IN SPOIIR-GLYC INTERGENIC REGION		isomerase	1025

938	80432586 (15031, 15032)	Novel Protein sim. GBank gil3202046 (AF069525) - 190 kDa ankyrin isoform; AnkG190 [Rattus norvegicus]	Contains protein domain (PF00023) - Ank repeat	kinase	1004, 1038
939	95292724 (13247, 13248)	Novel Protein sim. GBank gil4589628[db]BA76836.1I - (AB023209) KIAA0982 protein [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	kinase	1022
940	57293037 (14381, 14382)	Novel Protein sim. GBank gil115151[sp]P00546 CC28_YEAST - CELL DIVISION CONTROL PROTEIN 28	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1004, 1006, 1017, 1022, 1024, 1030, 1034, 1039
941	95010578 (991, 992)	Novel Protein sim. GBank gil1168818[sp]P06243 CC7_YEAST - CELL DIVISION CONTROL PROTEIN 7	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1017, 1024
942	78788409 (3619, 3620)	Novel Protein sim. GBank gil1170647[sp]P43637 KGS9_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE_YGL179C	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1022
943	33454872 (5145, 5146)	Novel Protein sim. GBank gil1228927 (L39263) - serine threonine protein kinase [Candida albicans]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1000, 1029
944	27842914 (5009, 5010)	Novel Protein sim. GBank gil1252591[sp]P15790 KC21_YEAST - CASEIN KINASE II, ALPHA CHAIN (CK II)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1010
945	78366860 (9391, 9392)	Novel Protein sim. GBank gil1257161[sp]P14681 KSS1_YEAST - MITOGEN- ACTIVATED PROTEIN KINASE KSS1 (MAP KINASE KSS1) (KINASE SUPPRESSOR OF SST2)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1008, 1009
946	27977744 (11365, 11366)	Novel Protein sim. GBank gil126820[sp]P21965 MCK1_YEAST - PROTEIN KINASE MCK1 (MEIOSIS AND CENTROMERE REGULATORY KINASE)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1006
947	39564971 (11909, 11910)	Novel Protein sim. GBank gil1330328 (U50595) - Rab8-interacting protein [Mus musculus]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1017, 1022, 1024

948	80240826 (1535, 1536)	Novel Protein sim. GBank gll134581sp P06782 SNF1_YEAST - CARBON CATABOLITE DERPRESSING PROTEIN KINASE	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1004, 1006, 1008, 1025, 1034
949	27841695 (14135, 14136)	Novel Protein sim. GBank gll1346383 sp P47116 PTK2_YEAST - SERINE/THREONINE-PROTEIN KINASE PTK2/STK2	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1010, 1024, 1027
950	99233916 (1831, 1832)	Novel Protein sim. GBank gll1404361 sp P23291 CK11_YEAST - CASEIN KINASE I HOMOLOG 1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1017, 1022, 1023, 1024, 1025, 1026, 1027, 1030, 1031, 1034, 1038, 1039, 1044, 1049
951	80233502 (3873, 3874)	Novel Protein sim. GBank gll1404361 sp P23291 CK11_YEAST - CASEIN KINASE I HOMOLOG 1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1006, 1008, 1010, 1017, 1022, 1034
952	17584851 (14635, 14636)	Novel Protein sim. GBank gll1586940 prfl [2205248A - Ser/Thr kinase] [Lycopersicon esculentum]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1003
953	80077096 (9483, 9484)	Novel Protein sim. GBank gll1711543 sp P50529 SSP1_SCHPO - SERINE/THREONINE-PROTEIN KINASE SSP1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1034
954	80085562 (7271, 7272)	Novel Protein sim. GBank gll177378 (M34146) - cell cycle protein (DBF2) [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1054

955	85519788 (12669, 12670)	Novel Protein sim. GBank gil1730060 spl P53894 KNQ1 - YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YNL161W	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1026, 1027, 1031, 1034, 1038, 1039, 1044, 1049, 1054
956	78730713 (7187, 7188)	Novel Protein sim. GBank gil173013 (M17074) - cAMP-dependent protein kinase subunit (put.); putative [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1008, 1022
957	39516001 (7573, 7574)	Novel Protein sim. GBank gil2132208 pir S65161 - hypothetical protein YPL150w - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1006, 1022, 1024, 1034
958	94668818 (13963, 13964)	Novel Protein sim. GBank gil2132208 pir S65161 - hypothetical protein YPL150w - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1022
959	79861968 (13115, 13116)	Novel Protein sim. GBank gil2132283 pir S59771 - hypothetical protein YPR106w - yeast [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1004, 1038
960	78381774 (14223, 14224)	Novel Protein sim. GBank gil2168137 emb CAA66149 - (X97547) PKF1 - [Fagus sylvatica]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1023
961	39529983 (14757, 14758)	Novel Protein sim. GBank gil2499619 spl Q03656 KM65 - PROBABLE SERINE/THREONINE-PROTEIN KINASE YMR216C	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1010
962	79186804 (22169, 22170)	Novel Protein sim. GBank gil2499621 spl Q12222 KOM8 - PROBABLE SERINE/THREONINE-PROTEIN KINASE YOL128C	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1006
963	10192971 (4731, 4732)	Novel Protein sim. GBank gil2654187 (AF034925) - MEKK-related kinase [Leishmania major]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1003

964	27802933 (22643, 22644)	Novel Protein sim. GBank gil295932[emb CAA68773] - (Y00867) PHO85 [Saccharomyces cerevisiae]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1024, 1034
965	78489683 (19937, 19938)	Novel Protein sim. GBank gil299627[bb5 129891 - Mp1p=mitogen-activated protein kinase homolog [Saccharomyces cerevisiae, Peptid, 484 aa]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1022
966	79594751 (14245, 14246)	Novel Protein sim. GBank gil30240531[sp Q82761 KC11_RAT - CASEIN KINASE I, GAMMA 1 ISOFORM (CK1-GAMMA 1)]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1008
967	71810946 (16165, 16166)	Novel Protein sim. GBank gil31231951[sp P32328 DBFB_YEAST - PROTEIN KINASE DBF20]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1008, 1022, 1038
968	3957659 (21331, 21332)	Novel Protein sim. GBank gil3367718[emb CAA06699] - (AJ005790) PKL12 [protein [Mus musculus]]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1006
969	23299249 (11877, 11878)	Novel Protein sim. GBank gil3643611 (AC005395) - putative casein kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1044
970	5496718 (4145, 4146)	Novel Protein sim. GBank gil39157451[sp P13186 KIN2_YEAST - PROTEIN KINASE KIN2]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1058
971	87452017 (17607, 17608)	Novel Protein sim. GBank gil45103421[gb AAD21431_1] - (AC006921) putative serine/threonine protein kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1025
972	79462327 (20037, 20038)	Novel Protein sim. GBank gil4512667[gb AAD21721_1] - (AC006931) putative MAP kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1034
973	80062951 (1829, 1830)	Novel Protein sim. GBank gil5102800[emb CAB45215_1] - (AL079308) putative serine/threonine protein kinase [Streptomyces coelicolor]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1024

974	10885459 (17451, 17452)	Novel Protein sim. GBank gi 510200 gb CA4B45215.1 - (AL079308) putative serine/threonine protein kinase [Streptomyces coelicolor]	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1024
975	78736707 (16853, 16854)	Novel Protein sim. GBank gi 532798 (U13398) - protein kinase [Saccharomyces cerevisiae]	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1044
976	21636169 (20485, 20486)	Novel Protein sim. GBank gi 53600681 gb AAD42851.1 (AF159689) serine/threonine kinase PKN3 [Myxococcus xanthus]	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1010, 1022
977	27369427 (3909, 3910)	Novel Protein sim. GBank gi 542984 pir JC2129 - protein kinase (EC 2.7.1.37) PKN - human	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1009
978	95007794 (15865, 15866)	Novel Protein sim. GBank gi 547783 sp P36005 KKQ1 YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL161C	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1017, 1023, 1039, 1044
979	11612738 (10929, 10930)	Novel Protein sim. GBank gi 585347 sp P38070 KBN8 YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YBR028C	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1004, 1024
980	11085515 (12089, 12090)	Novel Protein sim. GBank gi 585347 sp P38070 KBN8 YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YBR028C	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1004
981	20378278 (3053, 3054)	Novel Protein sim. GBank gi 588472 sp P36615 MDS1 YEAST - SERINE/THREONINE-PROTEIN KINASE MDS1/RIM11	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1024
982	27366497 (1059, 1060)	Novel Protein sim. GBank gi 732944 emb CAA57879 - (X82499) protein kinase [Saccharomyces cerevisiae]	Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1022
983	17299246 (3595, 3596)		Contains protein domain (PF000069) - Eukaryotic protein kinase domain	kinase	1044

984	10347705 (18865, 18866)		Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	1038
985	79837895 (11369, 11370)	Novel Protein sim. GBank gj 1653309 dbj BAA18224 - (D90912) regulatory components of sensory transduction system [Synechocystis sp.]	Contains protein domain (PF00072) - Response regulator receiver domain	kinase	1003, 1008
986	80494055 (11361, 11362)	Novel Protein sim. GBank gj 730745 sp P3992 SLN1_YEAST - OSOMOLARITY TWO-COMPONENT SYSTEM PROTEIN SLN1	Contains protein domain (PF00072) - Response regulator receiver domain	kinase	1006, 1009, 1010, 1012, 1022, 1024, 1034
987	27835820 (9257, 9258)	Novel Protein sim. GBank gj 172177 (M32491) - protein kinase C-like protein (PKC1) [Saccharomyces cerevisiae]	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	1006
988	79246868 (12621, 12622)	Novel Protein sim. GBank gj 172177 (M32491) - protein kinase C-like protein (PKC1) [Saccharomyces cerevisiae]	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	1008, 1009, 1011, 1022, 1026, 1039, 1044
989	8490496 (5415, 5416)	Novel Protein sim. GBank gj 1730510 sp P50319 PGKC_ALCEU - PHOSPHOGLYCERATE KINASE, CHROMOSOMAL	Contains protein domain (PF00162) - Phosphoglycerate kinases	kinase	1004
990	28389505 (15789, 15790)	Novel Protein sim. GBank gj 3738261 dbj BAA33803 - (AB018412) chloroplast phosphoglycerate kinase [Populus nigra]	Contains protein domain (PF00162) - Phosphoglycerate kinases	kinase	1044
991	94631242 (6675, 6676)	Novel Protein sim. GBank gj 4468682 emb CAB38136.1 - (AL035591) phosphoglycerate kinase [Streptomyces coelicolor]	Contains protein domain (PF00162) - Phosphoglycerate kinases	kinase	1008, 1009, 1010, 1024, 1031, 1044
992	30659716 (15717, 15718)	Novel Protein sim. GBank gj 4468682 emb CAB38136.1 - (AL035591) phosphoglycerate kinase [Streptomyces coelicolor]	Contains protein domain (PF00162) - Phosphoglycerate kinases	kinase	1022

993	80081347 (8333, 8334)	Novel Protein sim. GBank g 2132879 pir S61647 - probable membrane protein YOR086c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00168) - C2 domain	kinase	1003, 1034
994	79574330 (255, 256)	Novel Protein sim. GBank g 2497531 sp Q46078 KPYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	1031
995	99517532 (4255, 4256)	Novel Protein sim. GBank g 2497531 sp Q46078 KPYK_CORGL - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	1006, 1012, 1022, 1024, 1031, 1034
996	79326308 (4805, 4806)	Novel Protein sim. GBank g 3122312 sp O06134 KPYK_MYCTU - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	1034
997	80435060 (5213, 5214)	Novel Protein sim. GBank g 1172869 sp P4433 IRBSK_HAEIN - RIBOKINASE	Contains protein domain (PF00294) - pfkB family carbohydrate kinase	kinase	1023, 1038
998	79876170 (12305, 12306)	Novel Protein sim. GBank g 1172869 sp P4433 IRBSK_HAEIN - RIBOKINASE	Contains protein domain (PF00294) - pfkB family carbohydrate kinase	kinase	1017
999	71214416 (17373, 17374)	Novel Protein sim. GBank g 2132112 pir S67185 - hypothetical protein YOR283w - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00300) - Phosphoglycerate mutase family	kinase	1029
1000	78835506 (3541, 3542)	Novel Protein sim. GBank g 391414 sp P71904 NDK_MYCTU - NUCLEOSIDE DIPHOSPHATE KINASE (NDK) (NDP KINASE)	Contains protein domain (PF00334) - Nucleoside diphosphate kinases	kinase	1003
1001	79619378 (653, 854)	Novel Protein sim. GBank g 549688 sp P36089 YKG6_YEAST - HYPOTHETICAL 16.7 KD PROTEIN IN NDK1-MNR2 INTERGENIC REGION	Contains protein domain (PF00334) - Nucleoside diphosphate kinases	kinase	1006, 1008
1002	29273002 (2229, 2230)	Novel Protein sim. GBank g 125121 sp P16861 K6P1_YEAST - 6-PHOSPHOFRUCTOKINASE ALPHA SUBUNIT (PHOSPHOFRUCTOKINASE 1) (PHOSPHOHEXOKINASE) (6PF-1-K ALPHA SUBUNIT)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	1022

1003	78760314 (4001, 4002)	Novel Protein sim. GBank gi 1346355 sp Q01813 K6PP_HUMAN - 6-PHOSPHOFRUCTOKINASE, TYPE C (PHOSPHOFRUCTOKINASE 1) (PHOSPHOHEXOKINASE) (PHOSPHOFRUCTO-1-KINASE ISOZYME C) (6-PHOSPHOFRUCTOKINASE, PLATELET TYPE)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	1017
1004	82448765 (9871, 9872)	Novel Protein sim. GBank gi 3122290 sp Q08333 K6PF_STRCO - 6-PHOSPHOFRUCTOKINASE (PHOSPHOHEXOKINASE) (PHOSPHOHEXOKINASE) (ATP-PFK)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	1001, 1012, 1022, 1024, 1025
1005	80215680 (5173, 5174)	Novel Protein sim. GBank gi 3122305 sp Q27778 K6PF_SCHMA - 6-PHOSPHOFRUCTOKINASE (PHOSPHOHEXOKINASE) (PHOSPHOHEXOKINASE)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	1003, 1010, 1017, 1022, 1031
1006	95292832 (3795, 3796)	Novel Protein sim. GBank gi 5690378 gb AAD47060.1 AF_15950 - (AF_15950_1) Prk1 [Mycobacterium xanthus]	Contains protein domain (PF00365) - Phosphofructokinase	kinase	1006, 1010, 1024
1007	20750841 (735, 736)	Novel Protein sim. GBank gi 2493482 sp Q51390 GLPK_PSEAE - GLYCEROL KINASE (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	kinase	1022
1008	27842304 (11693, 11694)	Novel Protein sim. GBank gi 417058 sp P32190 GLPK_YEAST - GLYCEROL KINASE (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	kinase	1010
1009	81155814 (14289, 14290)	Novel Protein sim. GBank gi 417068 sp P32190 GLPK_YEAST - GLYCEROL KINASE (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK)	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	kinase	1006, 1017, 1024, 1026, 1054

1010	71781336 (15443, 15444)	Novel Protein sim. GBank gi 4980782 gb AAD35372.1 AE00171 - (AE00171/0) sugar kinase, FGGY family [Thermotoga maritima]	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	kinase	1026, 1029, 1036, 1053
1011	80216574 (819, 820)	Novel Protein sim. GBank gi 3122636 sp Q12417 PRL1_YEAST - PRL1/PRL2-LIKE PROTEIN	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1004, 1006, 1010, 1022, 1024, 1034
1012	9858201 (5825, 5826)	Novel Protein sim. GBank gi 4758560 ref NP_004805.1 pHPRP - U5 snRNP-specific 40 kDa protein (hPrp8-binding)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1023
1013	27923391 (9695, 9696)	Novel Protein sim. GBank gi 544075 sp P35605 COPP_BOVIN - COATOMER BETA' SUBUNIT (BETA'-COAT PROTEIN) (BETA'-COP) (P102)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1006
1014	80062828 (12959, 12960)	Novel Protein sim. GBank gi 549607 sp P36130 YK16_YEAST - HYPOTHETICAL 74.7 KD TRP-ASP REPEATS CONTAINING PROTEIN IN DAL80-GAP1 INTERGENIC REGION	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1008, 1017, 1024
1015	17708293 (8097, 8098)	Novel Protein sim. GBank gi 633348 emb CAA58712 - (X83754) alpha-COP (Saccharomyces cerevisiae)	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1008
1016	80586755 (13839, 13840)	Novel Protein sim. GBank gi 730431 sp P25635 PWP2_YEAST - PERIODIC TRYPTOPHAN PROTEIN 2	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1003, 1004, 1006, 1008, 1010, 1022, 1023, 1024, 1026, 1034, 1058
1017	99411986 (13153, 13154)	Novel Protein sim. GBank gi 731935 sp P40362 YJG9_YEAST - HYPOTHETICAL 66.4 KD PROTEIN IN SMC3-MRPL8 INTERGENIC REGION	Contains protein domain (PF00400) - WD domain, G-beta repeat	kinase	1003, 1004, 1006, 1008, 1010, 1012, 1016, 1022, 1024, 1034, 1044, 1054
1018	20634028 (22961, 22962)	Novel Protein sim. GBank gi 125155 sp P26364 KAD2_YEAST - ADENYLATE KINASE 2 (ATP-AMP TRANSPHOSPHORYLASE)	Contains protein domain (PF00406) - Adenylate kinase	kinase	1004

1019	11760014 (1567, 1568)	Novel Protein sim. GBank gj 2911007 emb CAA17500] - (AL021958) adk [Mycobacterium tuberculosis]	Contains protein domain (PF00406) - Adenylylate kinase	kinase	1022
1020	20416286 (11311, 11312)	Novel Protein sim. GBank gj 422060 pir A46079 - protein kinase pck2 (EC 2.7.1.-) - fission yeast (Schizosaccharomyces pombe)	Contains protein domain (PF00433) - Protein kinase C terminal domain	kinase	1004
1021	80249183 (1971, 1972)	Novel Protein sim. GBank gj 688242 (U77780) - secretory protein kinase [Chlorobium limicola]	Contains protein domain (PF00437) - Bacterial type II secretion system protein	kinase	1026, 1053
1022	65448379 (409, 410)	Novel Protein sim. GBank gj 608567 (U18116) - Bdf1p [Saccharomyces cerevisiae]	Contains protein domain (PF00439) - Bromodomain	kinase	1009
1023	79183453 (16947, 16948)	Novel Protein sim. GBank gj 139855 pir P16557 XYLR_BACSU - XYLOSE REPRESSOR	Contains protein domain (PF00480) - ROK family	kinase	1024, 1026, 1053
1024	91213694 (9093, 9094)	Novel Protein sim. GBank gj 2052193 emb CAA62003] - (X89964) glucokinase [Renibacterium salmoninarum]	Contains protein domain (PF00480) - ROK family	kinase	1004, 1025
1025	71095570 (11069, 11070)	Novel Protein sim. GBank gj 4160311 emb CAA22782] - (AL035212) putative ROK family transcriptional regulator [Streptomyces coelicolor]	Contains protein domain (PF00480) - ROK family	kinase	1053
1026	78719727 (18815, 18816)	Novel Protein sim. GBank gj 1710635 pir P51862 ROM2_YEAST - RHO1 GDP-GTP EXCHANGE PROTEIN 2	Contains protein domain (PF00610) - Domain found in Dishevelled, Egl-10, and Pleckstrin	kinase	1022, 1026
1027	79458038 (16501, 16502)	Novel Protein sim. GBank gj 2337797 emb CAA74271] - (Y13937) putative Gmk protein [Bacillus subtilis]	Contains protein domain (PF00625) - Guanylate kinase	kinase	1031
1028	77572881 (19547, 19548)	Novel Protein sim. GBank gj 1361215 pir S86581 - methyl-accepting chemotaxis protein I - Escherichia coli	Contains protein domain (PF00672) - Domain found in bacterial signal proteins	kinase	1023
1029	80059243 (4817, 4818)	Novel Protein sim. GBank gj 172241 (M85293) - glutamyl kinase [Saccharomyces cerevisiae]	Contains protein domain (PF00696) - Amino acid kinase family	kinase	1017, 1022, 1044

1030	79318332 (11575, 11576)	Novel Protein sim. GBank gi 1653564 dbj BAA18497 - (D90914) hypothetical protein [Synechocystis sp.]	Contains protein domain (PF00785) - PAC motif	kinase	1001, 1022
1031	79578021 (581, 582)	Novel Protein sim. GBank gi 4545127 gb AAD22405.1 AF07999 - (AF079997) aerotaxis receptor Aer [Pseudomonas putida]	Contains protein domain (PF00989) - PAS domain	kinase	1006
1032	80432645 (8761, 8762)	Novel Protein sim. GBank gi 1172627 sp P46546 PROB_CORGL - GLUTAMATE 5-KINASE (GAMMA-GLUTAMYL KINASE) (GK)	Contains protein domain (PF01472) - PUA domain	kinase	1004, 1006, 1008, 1010, 1022, 1026, 1034, 1038, 1054
1033	65446449 (10303, 10304)	Novel Protein sim. GBank gi 2182998 (U81489) - histidine kinase [Lactococcus lactis cremoris]	Contains protein domain (PF01569) - PAp2 superfamily	kinase	1016
1034	66398172 (17573, 17574)	Novel Protein sim. GBank gi 1168399 sp P41403 AKAB_MYCSM - ASPARTOKINASE ALPHA AND BETA SUBUNITS (ASPARTATE KINASE)	Contains protein domain (PF01842) - ACT domain	kinase	1008
1035	16771961 (5943, 5944)	Novel Protein sim. GBank gi 10027778 (U33326) - histidine protein kinase sensor CorS [Pseudomonas syringae pv. glycinea]		kinase	1008
1036	71757022 (3923, 3924)	Novel Protein sim. GBank gi 100796 pir S16585 - phosphomuokinase (EC 2.7.1.9) - wheat		kinase	1003, 1050
1037	11417809 (1399, 1400)	Novel Protein sim. GBank gi 1136667 sp P23960 ALUB_HUMAN - !!! ALU CLASS B WARNING ENTRY !!!		kinase	1017
1038	46572041 (15835, 15836)	Novel Protein sim. GBank gi 166399 sp P41403 AKAB_MYCSM - ASPARTOKINASE ALPHA AND BETA SUBUNITS (ASPARTATE KINASE)		kinase	1029
1039	78510452 (16611, 16612)	Novel Protein sim. GBank gi 1166399 sp P41403 AKAB_MYCSM - ASPARTOKINASE ALPHA AND BETA SUBUNITS (ASPARTATE KINASE)		kinase	1008

1040	78946031 (17791, 17792)	Novel Protein sim. GBank gil1171250 (U435686) - protein kinase related to Raf protein kinases; Method: conceptual translation supplied by author [Mus musculus]		kinase	1038, 1054
1041	95011545 (7529, 7530)	Novel Protein sim. GBank gil1175554 sp P44690 PDXY_HAEIN - PYRIDOXAMINE KINASE (PM KINASE)		kinase	1001
1042	17933518 (18521, 18522)	Novel Protein sim. GBank gil1176152 sp P44507 YHAD_HAEIN - HYPOTHETICAL PROTEIN H10091		kinase	1008
1043	79821849 (3881, 3882)	Novel Protein sim. GBank gil118198 sp P13510 CZCB_ALCEU - COBALT-T- ZINC-CADMIUM RESISTANCE PROTEIN CZCB (CATION EFFLUX SYSTEM PROTEIN CZCB)		kinase	1023
1044	79418477 (7175, 7176)	Novel Protein sim. GBank gil121064 sp P15442 GCN2_YEAST - PROTEIN KINASE GCN2		kinase	1004, 1024
1045	78770765 (9243, 9244)	Novel Protein sim. GBank gil1314785 (U55600) - protein kinase C suppressor Skc1p [Saccharomyces cerevisiae]		kinase	1008
1046	80496463 (2575, 2576)	Novel Protein sim. GBank gil134588 sp P06782 SNF1_YEAST - CARBON CATABOLITE DEREPRESSING PROTEIN KINASE		kinase	1012, 1022, 1024, 1034
1047	10324981 (8597, 8598)	Novel Protein sim. GBank gil134588 sp P06782 SNF1_YEAST - CARBON CATABOLITE DEREPRESSING PROTEIN KINASE		kinase	1025
1048	80502419 (1841, 1842)	Novel Protein sim. GBank gil1346383 sp P47116 PTK2_YEAST - SERINE/THREONINE-PROTEIN KINASE PTK2/STK2		kinase	1004, 1006, 1010, 1012, 1017, 1022, 1024

1049	20622000 (8419, 8420)	Novel Protein sim. GBank gil134837splP175801SPOT_ECOLI - GUANOSINE-3'-5'-BIS(DIPHOSPHATE) 3'-PYROPHOSPHOHYDROLASE ((PPGPP)ASE) (PENTA-PHOSPHATE GUANOSINE-3'-PYROPHOSPHOHYDROLASE)		kinase	1004
1050	20620598 (12031, 12032)	Novel Protein sim. GBank gil140201splP146801YAK1_YEAST - PROTEIN KINASE YAK1		kinase	1022
1051	78677965 (15499, 15500)	Novel Protein sim. GBank gil140436splP2329ICK11_YEAST - CASEIN KINASE I HOMOLOG 1		kinase	1009
1052	65903201 (22259, 22260)	Novel Protein sim. GBank gil154289splembICAB021931 - (Z80108) mlf [Mycobacterium tuberculosis]		kinase	1023
1053	20434454 (14773, 14774)	Novel Protein sim. GBank gil1542971 (U50713) - AlgZ [Pseudomonas aeruginosa]		kinase	1022
1054	80502165 (7301, 7302)	Novel Protein sim. GBank gil1652132 dbj BAA17056 - (D90903) sensory transduction histidine kinase [Synechocystis sp.]		kinase	1012
1055	66490420 (13409, 13410)	Novel Protein sim. GBank gil1652639 dbj BAA17559 - (D90907) sensory transduction histidine kinase [Synechocystis sp.]		kinase	1029
1056	10265715 (4967, 4968)	Novel Protein sim. GBank gil1653075 dbj BAA17992 - (D90910) sensory transduction histidine kinase [Synechocystis sp.]		kinase	1044
1057	25148357 (16461, 16462)	Novel Protein sim. GBank gil1688242 (U77780) - secretory protein kinase [Chlorobium limicola]		kinase	1026
1058	27999872 (17293, 17294)	Novel Protein sim. GBank gil1694854 emb CAB05435 - (Z83018) thiL [Mycobacterium tuberculosis]		kinase	1044

1059	8489622 (1527, 1528)	Novel Protein sim. GBank gil1706182 sp P46963 CTK3 YEAST - CTD KINASE GAMMA SUBUNIT (CTD KINASE 32 KD SUBUNIT) (CTDK-I GAMMA SUBUNIT)		kinase	1034
1060	78360895 (22797, 22798)	Novel Protein sim. GBank gil1711083 sp P52560 RELA_STRCO - GTP PYROPHOSPHOKINASE (ATP:GTP 3'-PYROPHOSPHOTRANSFERASE) (PPGPP SYNTHETASE I)		kinase	1008, 1026, 1030
1061	95292907 (19927, 19928)	Novel Protein sim. GBank gil171204 (K01763) - CDC8 gene [Saccharomyces cerevisiae]		kinase	1010
1062	25323314 (141, 142)	Novel Protein sim. GBank gil171789 (M69018) - protein kinase 2 [Saccharomyces cerevisiae]		kinase	1017
1063	16297568 (17091, 17092)	Novel Protein sim. GBank gil1729784 emb CA70932 - (Y09798) colS [Pseudomonas fluorescens]		kinase	1054
1064	80221856 (4985, 4986)	Novel Protein sim. GBank gil1730671 sp P53727 YN8F YEAST - HYPOTHETICAL 35.4 KD PROTEIN IN SEC12-SSK2 INTERGENIC REGION		kinase	1004, 1008, 1023
1065	11691238 (3659, 3660)	Novel Protein sim. GBank gil1731204 sp P54833 YY16_MYCLE - PROBABLE SENSOR-LIKE_HISTIDINE KINASE B2168_C3_247		kinase	1034
1066	80217625 (201, 202)	Novel Protein sim. GBank gil173179 (M59835) - protein kinase [Saccharomyces cerevisiae]		kinase	1004, 1006, 1022, 1023, 1024, 1025, 1026, 1027, 1038
1067	79588225 (18895, 18896)	Novel Protein sim. GBank gil173179 (M59835) - protein kinase [Saccharomyces cerevisiae]		kinase	1010, 1026, 1034
1068	87114784 (3707, 3708)	Novel Protein sim. GBank gil1787450 (AE000218) - putative dihydroxyacetone kinase (EC 2.7.1.2) [Escherichia coli]		kinase	1038
1069	87464582 (16879, 16880)	Novel Protein sim. GBank gil1834376 emb CAA70860 - (Y09666) RagB [Bradyrhizobium japonicum]		kinase	1048

1070	79873185 (7115, 7116)	Novel Protein sim. GBank gii 1839006 emb [CAB06648] - (Z85982) argB [Mycobacterium tuberculosis]		kinase	1001, 1017, 1026, 1049
1071	36991618 (17217, 17218)	Novel Protein sim. GBank gii 1869996 emb [CAB06847] - (Z92539) hypothetical protein Rv1032c [Mycobacterium tuberculosis]		kinase	1016
1072	28846324 (21259, 21260)	Novel Protein sim. GBank gii 2072305 (J95145) - S-AKAP81 [Mus musculus]		kinase	1026
1073	85439100 (12739, 12740)	Novel Protein sim. GBank gii 2076894 gb JAB63983.1 - (AE002197) short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG_PE-bind), Score=10.0, E-value=0.0034, N=1 [Caenorhabditis elegans]		kinase	1006, 1038
1074	8756691 (10079, 10080)	Novel Protein sim. GBank gii 2131279 pir S50157 - cyclin-dependent kinase chain SRB10 - yeast (Saccharomyces cerevisiae)		kinase	1004
1075	20612864 (7583, 7584)	Novel Protein sim. GBank gii 218490 dbj BA01226 - (D10389) Ssp31 protein kinase [Saccharomyces cerevisiae]		kinase	1022
1076	29348929 (7611, 7612)	Novel Protein sim. GBank gii 2251187 emb [CAB10669.1 - (Z97369) uridylylate kinase [Mycobacterium leprae]		kinase	1022
1077	79856633 (19069, 19070)	Novel Protein sim. GBank gii 226292 pir 1505375A - vir gene [Bordetella pertussis]		kinase	1017
1078	80209820 (3877, 3878)	Novel Protein sim. GBank gii 231338 gb AAD07313.1 - (AE000544) signal- transducing protein, histidine kinase (atoS) [Helicobacter pylori] 26695		kinase	1034
1079	94327239 (5447, 5448)	Novel Protein sim. GBank gii 231831 sp P29931 COBP_PSEDE - COBINAMIDE KINASE / COBINAMIDE PHOSPHATE GUANYLYLTRANSFERASE		kinase	1010

			kinase	1006
1080	11090638 (17233, 17234)	Novel Protein sim. GBank gi 2462047 emb CAA86935.1 - (Z46863) polyphosphate kinase [Acinetobacter sp. ADP1]		
1081	27849196 (6653, 6654)	Novel Protein sim. GBank gi 2462749 (AC002292) - Putative Serine/Threonine protein kinase [Arabidopsis thaliana]	kinase	1006, 1022
1082	16328294 (19421, 19422)	Novel Protein sim. GBank gi 2462827 (AF000657) - probable thiamin biosynthetic enzyme [Arabidopsis thaliana]	kinase	1034
1083	20615747 (4523, 4524)	Novel Protein sim. GBank gi 2463029 emb CAA74020 - (Y13670) sensory kinase [Erwinia carotovora subsp. carotovora]	kinase	1010
1084	65641884 (11511, 11512)	Novel Protein sim. GBank gi 2492975 spi P95014 AROK_MYTCTU - SHIKIMATE KINASE (SK)	kinase	1016
1085	70957866 (8397, 8398)	Novel Protein sim. GBank gi 2498150 spi P94524 ARAB_BACSU - L- RIBULOKEINASE	kinase	1029
1086	80048433 (13275, 13276)	Novel Protein sim. GBank gi 2489003 spi P76422 THID_ECOLI - PHOSPHOMETHYL PYRIMIDINE KINASE (HMP- PHOSPHATE KINASE) (HMP-P KINASE)	kinase	1034
1087	11690337 (8693, 8694)	Novel Protein sim. GBank gi 2489009 spi Q57190 THIL_HAEIN - THIAMIN- MONOPHOSPHATE KINASE (THIAMIN- PHOSPHATE KINASE)	kinase	
1088	87462116 (13123, 13124)	Novel Protein sim. GBank gi 2499489 spi Q41141 PFPB_RICCO - PYROPHOSPHATE-6-FRUCTOSE 6- PHOSPHATE 1-PHOSPHOTRANSFERASE BETA SUBUNIT (PFP) (6- PHOSPHOFRUCTOKINASE (PYROPHOSPHATE)) (PYROPHOSPHATE- DEPENDENT 6-PHOSPHOFRUCTOSE-1- KINASE) (PPI-PFK)	kinase	1025

1089	80582001 (1513, 1514)	Novel Protein sim. GBank gi 2499623 sp Q12236 KOK0_ YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YOL100W		kinase	1010, 1022, 1034, 1058
1090	8485449 (3809, 3810)	Novel Protein sim. GBank gi 2499629 sp Q6085 RIP_MOUSE - SERINE/THREONINE PROTEIN KINASE RIP (CELL DEATH PROTEIN RIP) (RECEPTOR INTERACTING PROTEIN)		kinase	1022
1091	57280428 (22145, 22146)	Novel Protein sim. GBank gi 2507190 sp P22517 KCC2_ YEAST - CALMODULIN-DEPENDENT PROTEIN KINASE II		kinase	1058
1092	80402775 (20139, 20140)	Novel Protein sim. GBank gi 2555172 (AF025543) - ArcC; carbamate kinase [Rhizobium etii]		kinase	1006, 1022, 1024, 1029, 1034
1093	799633862 (15695, 15696)	Novel Protein sim. GBank gi 2580433 dbj BAA23138 - (D764-14) ppGpp hydrolase [Staphylococcus aureus]		kinase	1022
1094	99447225 (7879, 7880)	Novel Protein sim. GBank gi 2623815 (AF030352) - two component sensor [Pseudomonas aeruginosa]		kinase	1023, 1039
1095	27927751 (3017, 3018)	Novel Protein sim. GBank gi 2739052 (AF025310) - tsk-1 and tsk-2 kinase substrate [Mus musculus]		kinase	1006, 1010, 1024
1096	39436442 (5539, 5540)	Novel Protein sim. GBank gi 2739052 (AF025310) - tsk-1 and tsk-2 kinase substrate [Mus musculus]		kinase	1004, 1008, 1022, 1058
1097	28831811 (5435, 5436)	Novel Protein sim. GBank gi 2896729 emb CAA17208.1 - (AL021897) coaA [Mycobacterium tuberculosis]		kinase	1022
1098	85802999 (18555, 18556)	Novel Protein sim. GBank gi 2916943 emb CAA17581 - (AL021999) hypothetical protein Rv0982 [Mycobacterium tuberculosis]		kinase	1038

1099	79631651 (11411, 11412)	Novel Protein sim. GBank gi 2961066 (AF013273) - histidine kinase 1; CaHK1 [Candida albicans]		kinase	1025, 1044
1100	87453156 (8189, 8190)	Novel Protein sim. GBank gi 2961358 emb CAA18116.1 - (AL022140) serine/threonine protein kinase like protein [Arabidopsis thaliana]		kinase	1025
1101	78077910 (6093, 6094)	Novel Protein sim. GBank gi 29955292 emb CAA18321 - (AL022268) putative two-component sensor [Streptomyces coelicolor]		kinase	1029, 1053
1102	78675446 (383, 384)	Novel Protein sim. GBank gi 3025046 sp P75844 YCAR_ECOLI - HYPOTHETICAL 6.9 KD PROTEIN IN MSBA-KDSB INTERGENIC REGION		kinase	1008
1103	20439255 (22915, 22916)	Novel Protein sim. GBank gi 31222314 sp O34971 KDPD_RATRA - SENSOR PROTEIN KDPD		kinase	1024
1104	20464942 (17579, 17580)	Novel Protein sim. GBank gi 3150280 emb CAA19179 - (AL023634) cyclin [Schizosaccharomyces pombe]		kinase	1010
1105	10156682 (5439, 5440)	Novel Protein sim. GBank gi 3256535 dbj BAA29218.1 - (AP0000001) 301aa long hypothetical 2-phosphoglycerate kinase [Pyrococcus horikoshii]		kinase	1003
1106	11394864 (679, 680)	Novel Protein sim. GBank gi 3355803 1 dbj BAA31995 - (AB015974) glycerol kinase [Pseudomonas tlaaesi]		kinase	1006
1107	28461459 (2319, 2320)	Novel Protein sim. GBank gi 3522970 (U42390) - Trifl [Homo sapiens]		kinase	1003
1108	11011597 (6639, 6640)	Novel Protein sim. GBank gi 3702958 (AF077659) - homeodomain-interacting protein kinase 2 [Mus musculus]		kinase	1001
1109	79831334 (11591, 11592)	Novel Protein sim. GBank gi 4033487 sp Q44472 TUD4_AGRVI - PUTATIVE HYDROXYPYRUVATE REDUCTASE		kinase	1023

1110	10251094 (22969, 22970)	Novel Protein sim. G-Bank g 4 103624 (AF026470) -glucuronate kinase [Pseudomonas aeruginosa]		Kinase	1038
1111	78473540 (575, 576)	Novel Protein sim. G-Bank g 4 104603 (AF036966) -putative histidine kinase [Lactobacillus sakei]		Kinase	1026
1112	13042440 (8529, 8530)	Novel Protein sim. G-Bank g 4 17640 sp P32 17 RHAB_ECOLI - RHAMNULOKINASE (RHAMNULOSE PROTEIN KINASE)		Kinase	1024
1113	79430394 (6319, 6320)	Novel Protein sim. G-Bank g 4 17775 sp Q01389 BCK1_YEAST - SERINE/THREONINE PROTEIN KINASE BCK1 SLK1 SSP31		Kinase	1017
1114	78180866 (271, 272)	Novel Protein sim. G-Bank g 4 4642 sp emb CAB37472 - (AL035559) probable riboflavin kinase (FAD synthetase) [Streptomyces coelicolor]		Kinase	1053
1115	82124905 (10323, 10324)	Novel Protein sim. G-Bank g 4 5395 sp emb CAB38479_1 - (AL035636) Integral membrane protein with kinase activity [Streptomyces coelicolor]		Kinase	1001, 1004, 1016, 1024, 1031
1116	8367484 (5799, 5800)	Novel Protein sim. G-Bank g 4 5672 sp g AAD23692_1 AC00684 - (AC006841) putative serine/threonine protein - kinase [Arabidopsis thaliana]		Kinase	1006
1117	37797470 (2377, 2378)	Novel Protein sim. G-Bank g 4 62451 sp P34244 KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W		Kinase	1012
1118	78746126 (6789, 6790)	Novel Protein sim. G-Bank g 4 62451 sp P34244 KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W		Kinase	1017, 1022, 1026

1119	27364856 (13131, 13132)	Novel Protein sim. GBank g 46245 sp P34244 KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W		kinase	1009
1120	79831737 (21057, 21058)	Novel Protein sim. GBank g 46245 sp P34244 KKK1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL101W		kinase	1008, 1009, 1017
1121	20468860 (6679, 6680)	Novel Protein sim. GBank g 46889271 gb AAD27839_1 AF12460 - (AF12460) shikimate kinase [Corynebacterium glutamicum]		kinase	1010
1122	71456164 (16591, 16592)	Novel Protein sim. GBank g 479395 pir S33653 - probable serine/threonine protein kinase (EC 2.7.1.-) - yeast (Saccharomyces cerevisiae)		kinase	1022, 1026
1123	30519332 (12231, 12232)	Novel Protein sim. GBank g 4826712 ref NP_004945.1 pEMK1 - ELKL motif kinase		kinase	1024
1124	78497861 (20563, 20564)	Novel Protein sim. GBank g 4826712 ref NP_004945.1 pEMK1 - ELKL motif kinase		kinase	1038
1125	70945142 (18931, 18932)	Novel Protein sim. GBank g 4980554 gb AAD35161_1 AE001669 - (AE001669) 2-keto-3-deoxygluconate kinase [Thermotoga maritima]		kinase	1029
1126	79618379 (9063, 9064)	Novel Protein sim. GBank g 5019771 gb AAD37857_1 AF13326 - (AF13326) histidine protein kinase-response regulator hybrid protein CvgSY [Pseudomonas syringae pv. syringae]		kinase	1039
1127	94665374 (2841, 2842)	Novel Protein sim. GBank g 5114231 gb AAD40238_1 AF13670 - (AF13670) histidine kinase YcgG [Staphylococcus aureus]		kinase	1024

1128	79969348 (17053, 17054)	Novel Protein sim. GBank gi 5114231 gb AAD40238.1 AF13670 - (AF13670) histidine kinase YycG [Staphylococcus aureus]		kinase	1026, 1034
1129	19899327 (17811, 17812)	Novel Protein sim. GBank gi 5123787 emb CAg5385.1 - (AJ007041) trithorax homologue 2 [Homo sapiens]		kinase	1054
1130	78678510 (16361, 16362)	Novel Protein sim. GBank gi 5174547 ref NP_005914.1 pMEKK - MAP/ERK kinase kinase 5		kinase	1017
1131	11614075 (20663, 20664)	Novel Protein sim. GBank gi 532798 (U13398) - protein kinase [Saccharomyces cerevisiae]		kinase	1024
1132	25320879 (22705, 22706)	Novel Protein sim. GBank gi 547783 sp P36005 KKQ1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL161C		kinase	1009
1133	29350065 (8933, 8934)	Novel Protein sim. GBank gi 547784 sp P36004 KKQ8_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL168C		kinase	1009, 1022
1134	78780725 (16617, 16618)	Novel Protein sim. GBank gi 547785 sp P36003 KKR1_YEAST - PROBABLE SERINE/THREONINE-PROTEIN KINASE YKL171W		kinase	1009
1135	25266794 (1803, 1804)	Novel Protein sim. GBank gi 557484 (U15300) - Hal5p [Saccharomyces cerevisiae]		kinase	1008
1136	79863441 (14917, 14918)	Novel Protein sim. GBank gi 625679 pir A36929 - virulence regulatory protein VsrB - Pseudomonas solanacearum		kinase	1003
1137	28384835 (3043, 3044)	Novel Protein sim. GBank gi 828037 pir A533731 - translation initiation factor eIF-2 alpha chain kinase (EC 2.7.1.-) - rat		kinase	1008
1138	46866634 (5449, 5450)	Novel Protein sim. GBank gi 729831 sp P39188 ALU1_HUMAN - !!! ALU SUBFAMILY J WARNING ENTRY!!!		kinase	1037

1139	10238144 (22055, 22056)	Novel Protein sim. GBank gi 728831 sp P3918 ALU1_HUMAN - !!! ALU SUBFAMILY J WARNING ENTRY !!!		kinase	1023
1140	32552813 (22507, 22508)	Novel Protein sim. GBank gi 728831 sp P3918 ALU1_HUMAN - !!! ALU SUBFAMILY J WARNING ENTRY !!!		kinase	1006
1141	83335924 (13001, 13002)	Novel Protein sim. GBank gi 728835 sp P39192 ALU5_HUMAN - !!! ALU SUBFAMILY SC WARNING ENTRY !!!		kinase	1054
1142	28395057 (9957, 9958)	Novel Protein sim. GBank gi 728833 sp P39195 ALU8_HUMAN - !!! ALU SUBFAMILY SX WARNING ENTRY !!!		kinase	1025
1143	30507374 (21575, 21576)	Novel Protein sim. GBank gi 729900 sp Q01919 KIN4_YEAST - SERINE/THREONINE-PROTEIN KINASE KIN4		kinase	1009
1144	27980343 (827, 828)	Novel Protein sim. GBank gi 729911 sp P3869 JKSP1_YEAST - SERINE/THREONINE-PROTEIN KINASE KSP1		kinase	1022
1145	27850048 (18749, 18750)	Novel Protein sim. GBank gi 729911 sp P3869 JKSP1_YEAST - SERINE/THREONINE-PROTEIN KINASE KSP1		kinase	1034
1146	82337043 (14315, 14316)	Novel Protein sim. GBank gi 731646 sp P3869 YHJ3_YEAST - HYPOTHETICAL 46.9 KD PROTEIN IN SLT2- PUT2 INTERGENIC REGION		kinase	1001, 1004, 1006, 1008, 1009, 1010, 1017, 1022, 1023, 1024, 1034, 1038, 1044

1147	99283252 (22091, 22092)	Novel Protein sim. GBank gi 7316481 sp P38690 YHJ3 _YEAST - HYPOTHETICAL 46.9 KD PROTEIN IN SLT2- PUT2 INTERGENIC REGION		kinase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1013, 1016, 1017, 1018, 1022, 1023, 1024, 1025, 1026, 1027, 1029, 1031, 1034, 1038, 1039, 1044, 1050, 1054
1148	78767831 (4337, 4338)	Novel Protein sim. GBank gi 854061 emb CAA563371 - (X83413) U88 [Human herpesvirus 6]		kinase	1008
1149	20282021 (191, 192)			kinase	1010
1150	28461467 (3473, 3474)			kinase	1003
1151	29469151 (851, 852)	Novel Protein sim. GBank gi 1345537 sp P49106 1431_MAIZE - 14-3-3-LIKE PROTEIN GF14-6	Contains protein domain (PF00244) - 14-3-3 proteins	kinaseinhibitor	1026
1152	29462916 (2793, 2794)	Novel Protein sim. GBank gi 1345537 sp P49106 1431_MAIZE - 14-3-3-LIKE PROTEIN GF14-6	Contains protein domain (PF00244) - 14-3-3 proteins	kinaseinhibitor	1026
1153	23294238 (7325, 7326)	Novel Protein sim. GBank gi 2492411 sp Q99002 1433_TRIHA - 14-3-3 PROTEIN HOMOLOG (TH1433)	Contains protein domain (PF00244) - 14-3-3 proteins	kinaseinhibitor	1003
1154	79574588 (3941, 3942)	Novel Protein sim. GBank gi 2982876 (AE000675) - protein kinase C inhibitor (HIT family) [Aquifex aeolicus]	Contains protein domain (PF01230) - HIT family	kinaseinhibitor	1031
1155	52561371 (10137, 10138)	Novel Protein sim. GBank gi 3023940 sp O07513 HIT_BACSU - HIT PROTEIN	Contains protein domain (PF01230) - HIT family	kinaseinhibitor	1034

1156	37807912 (17343, 17344)	Novel Protein sim. GBank gi 2911080 emb CAA17542.1 - (AL021960) receptor kinase-like protein [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinasereceptor	1058
1157	29017363 (18983, 18984)	Novel Protein sim. GBank gi 5679843 emb CAB51836.1 - (AJ243961) Putative Ser/Thr protein kinase [Oryza sativa]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinasereceptor	1008
1158	99413939 (15555, 15556)	Novel Protein sim. GBank gi 13461091 sp P49027 GBLP_ORYSA - GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN (GPB-LR) (RWD)	Contains protein domain (PF00400) - WD domain, G- beta repeat	kinasereceptor	1009, 1010, 1034, 1044
1159	656687073 (839, 840)	Novel Protein sim. GBank gi 2392895 (AF07056) - brassinosteroid insensitive 1 [Arabidopsis thaliana]	Contains protein domain (PF00560) - Leucine Rich Repeat	kinasereceptor	1049
1160	78384656 (9399, 9400)	Novel Protein sim. GBank gi 3033389 (AC004238) - Cf-2.1-like protein [Arabidopsis thaliana]	Contains protein domain (PF00560) - Leucine Rich Repeat	kinasereceptor	1017
1161	37807938 (20131, 20132)	Novel Protein sim. GBank gi 453938 emb CAB37449.1 - (AL035526) putative protein (fragment) [Arabidopsis thaliana]	Contains protein domain (PF00560) - Leucine Rich Repeat	kinasereceptor	1058
1162	14994663 (8361, 8362)	Novel Protein sim. GBank gi 2497357 sp P55500 Y4 Q_RHISN - PUTATIVE INSERTION SEQUENCE ATP-BINDING PROTEIN Y4 Q_Y4ND Y4SD		kinasereceptor	1024
1163	10137798 (17309, 17310)	Novel Protein sim. GBank gi 1857771 gb AAB4482 - (U87224) contactin associated protein [Rattus norvegicus]	Contains protein domain (PF00054) - Laminin G domain	laminin	1039
1164	30575607 (3951, 3952)	Novel Protein sim. GBank gi 1586274 prt 2203365A - laminin alpha5 [Mus musculus]		laminin	1009
1165	20752912 (4855, 4856)	Novel Protein sim. GBank gi 1986274 prt 2203365A - laminin alpha5 [Mus musculus]		laminin	1022

1166	65663524 (2971, 2972)	Novel Protein sim. GBank gi 34492941 dbj BAA324621 - (AB011532) MEGF6 [Rattus norvegicus]		laminin	1053
1167	90938310 (12657, 12658)	Novel Protein sim. GBank gi 34492941 dbj BAA324621 - (AB011532) MEGF6 [Rattus norvegicus]		laminin	1020
1168	38915597 (7657, 7658)	Novel Protein sim. GBank gi 111209 pir A34671 - triacylglycerol lipase (EC 3.1.1.3) precursor, pancreatic - mouse		lipase	1026
1169	23307453 (16093, 16094)	Novel Protein sim. GBank gi 1430921 emb CAA67627 - (X99255) triacylglycerol lipase [Propionibacterium acnes]		lipase	1008
1170	10185115 (17103, 17104)	Novel Protein sim. GBank gi 1542917 emb CAB021801 - (Z80108) lipH [Mycobacterium tuberculosis]		lipase	1003
1171	350233738 (8389, 8390)	Novel Protein sim. GBank gi 1694831 emb CAB054411 - (Z83018) lipN [Mycobacterium tuberculosis]		lipase	1053
1172	8490153 (4515, 4516)	Novel Protein sim. GBank gi 1708847 sp P54310 LPS_MOUSE - HORMONE SENSITIVE LIPASE (HSL)		lipase	1022
1173	85801114 (2905, 2906)	Novel Protein sim. GBank gi 2555144 (AF024666) - transposase 5564 [Corynebacterium striatum]		lipase	1008, 1025, 1029
1174	39566187 (4509, 4510)	Novel Protein sim. GBank gi 5174497 ref NP_006024.1 pLIPG - lipase, endothelial		lipase	1006
1175	88094176 (341, 342)	Novel Protein sim. GBank gi 1155491 sp P27825 CALX_YEAST - CALNEXIN HOMOLOG PRECURSOR	Contains protein domain (PF00262) - Calreticulin family	MHC	1004
1176	80053908 (19709, 19710)	Novel Protein sim. GBank gi 1155491 sp P27825 CALX_YEAST - CALNEXIN HOMOLOG PRECURSOR	Contains protein domain (PF00262) - Calreticulin family	MHC	1003, 1008, 1009, 1017, 1022, 1023, 1024, 1026, 1034, 1044

1177	87114558 (13565, 13586)	Novel Protein sim. GBank gi 1181331 emb CAA54678] - (X77569) calnexin [Zea mays]	Contains protein domain (PF00262) - Calreticulin family	MHC	1025
1178	78799319 (4763, 4784)	Novel Protein sim. GBank gi 1181331 emb CAA54975] - (X78057) calreticulin [Zea mays]	Contains protein domain (PF00262) - Calreticulin family	MHC	1022
1179	78907081 (22469, 22470)	Novel Protein sim. GBank gi 117936 sp P14922 SSN6 YEAST - GLUCOSE REPRESSION MEDIATOR PROTEIN	Contains protein domain (PF00515) - TPR Domain	MHC	1003
1180	40388909 (15551, 15552)	Novel Protein sim. GBank gi 113667 sp P23960 ALUB_HUMAN - !!! ALU CLASS B WARNING ENTRY !!!		MHC	1037
1181	79261022 (1977, 1978)	Novel Protein sim. GBank gi 170871 emb CAA66313] - (X97729) LMP7-like protein [Botryllus schlosseri]		MHC	1003
1182	47652760 (2005, 2006)	Novel Protein sim. GBank gi 3589347 (AF081114) - ORF2 [Mus musculus domesticus]		MHC	1029
1183	78678292 (15127, 15128)	Novel Protein sim. GBank gi 549045 sp P36371 TAP2_MOUSE - ANTIGEN PEPTIDE TRANSPORTER 2 (APT2) (HISTOCOMPATIBILITY ANTIGEN MODIFIER 2)		MHC	1017
1184	17129704 (11867, 11868)	Novel Protein sim. GBank gi 553190 gb AAD44524.1 AF10740 - (AF10740) GW128 [Homo sapiens]		MHC	1038
1185	15024608 (16595, 16596)	Novel Protein sim. GBank gi 108693 pir A40437 - glutamic acid-rich protein, retinal - bovine	misc_channel	1024	

1186	82124914 (12671, 12672)	Novel Protein sim. GBank gi 2493749 sp Q28181 [CNG4_BOVIN - 240 KD PROTEIN OF ROD PHOTORECEPTOR CNG- CHANNEL [CONTAINS: GLUTAMIC ACID-RICH PROTEIN (GARP); CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4 (CNG CHANNEL 4) (CYCLIC NUCLEOTIDE-GATED CATION CHANNEL MODULATORY SUBUNIT)]		misc_channel	1016
1187	20297584 (20271, 20272)	Novel Protein sim. GBank gi 2916946 emb CAA17584 - (AL021999) mscl [Mycobacterium tuberculosis]		misc_channel	1034
1188	80069017 (14017, 14018)	Novel Protein sim. GBank gi 3242244 emb CAA12408 - (AJ225124) hyperpolarization-activated cation channel, HAC3 [Mus musculus]		misc_channel	1022, 1034
1189	82116763 (3257, 3258)	Novel Protein sim. GBank gi 11686818 sp P25558 [BUD3_YEAST - BUD SITE SELECTION PROTEIN BUD3 [Mus musculus]		nucl_recept	1003, 1006, 1008, 1016, 1017, 1022, 1023, 1024, 1044
1190	78377027 (17201, 17202)	Novel Protein sim. GBank gi 2529416 [U73935] - unknown [Shewanella sp. SCRC-2738]		nucl_recept	1026
1191	79610471 (18823, 18824)	Novel Protein sim. GBank gi 2529416 [U73935] - unknown [Shewanella sp. SCRC-2738]		nucl_recept	1004
1192	8910216 (9231, 9232)	Novel Protein sim. GBank gi 3378429 [AF079317] - unknown [Sphingomonas aromaticivorans]		nucl_recept	1023
1193	80538546 (7613, 7614)	Novel Protein sim. GBank gi 416584 sp Q04752 [AD4B_BOVIN - STEROID HORMONE RECEPTOR AD4BP (STEROIDOGENIC FACTOR 1) (STF-1) (SF-1) (FUSHI TARAZU FACTOR HOMOLOG 1)		nucl_recept	1004, 1022, 1024, 1058
1194	34894462 (8767, 8768)	Novel Protein sim. GBank gi 4887756 gb AAD32292.1 AC006533 - (AC006533) putative protein kinase [Arabidopsis thaliana]	Contains protein domain (PF000023) - Ank repeat	nuclease	1050

1195	9388465 (12705, 12706)	Novel Protein sim. GBank gil132231isp P11406 RECA_PROMI - RECA PROTEIN	Contains protein domain (PF00154) - recA bacterial DNA recombination proteins	nuclease	1016, 1017
1196	65984678 (7399, 7400)	Novel Protein sim. GBank gil2500086 isp Q59560 RECA_MYCSM - RECA PROTEIN	Contains protein domain (PF00154) - recA bacterial DNA recombination proteins	nuclease	1004
1197	36856857 (10825, 10826)	Novel Protein sim. GBank gil1174918 isp P45125 UVRB_HAEIN - EXCINUCLEASE ABC SUBUNIT B	Contains protein domain (PF00271) - recA bacterial DNA recombination proteins	nuclease	1016
1198	87457250 (14431, 14432)	Novel Protein sim. GBank gil458558 emb CA0855.1 - (AL049628) putative adenine glycosylase [Streptomyces coelicolor]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deoR family	nuclease	1001, 1004, 1006, 1010, 1012, 1022, 1024, 1025, 1026, 1034, 1039
1199	78365706 (8387, 8388)	Novel Protein sim. GBank gil3309125 (AF053745) - pol polyprotein [Mus dunni endogenous virus] [Mus dunni endogenous virus]	Contains protein domain (PF00552) - Integrase	nuclease	1003
1200	71095127 (21871, 21872)	Novel Protein sim. GBank gil400773 emb CAA22415 - (AL034447) ribonuclease III [Streptomyces coelicolor]	Contains protein domain (PF00636) - RNase3 domain.	nuclease	1053
1201	80496304 (8565, 8566)	Novel Protein sim. GBank gil2960098 emb CAA17996.1 - (AL022121) nth [Mycobacterium tuberculosis]	Contains protein domain (PF00730) - Endonuclease III	nuclease	1004, 1006, 1009, 1010, 1012, 1023, 1034
1202	95287851 (2175, 2176)	Novel Protein sim. GBank gil187736 emb CA057118 - (Z92772) recD [Mycobacterium tuberculosis]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	1004, 1012, 1022, 1024, 1034
1203	80079401 (19335, 19336)	Novel Protein sim. GBank gil3142727 gb AAC16891.1 - (AF064802) exodeoxyribonuclease V subunit [Photobacterium profundum]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	1034

1204	87913351 (17849, 17850)	Novel Protein sim. GBank gi 553133 emb CAB50985.1 - (AL096852), putative ribonuclease PH [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01726) - Ham1 family	nuclease	1003, 1008, 1015, 1022, 1025, 1026
1205	77520587 (17765, 17766)	Novel Protein sim. GBank gi 1023112 (U36840) - CG Site No. 10872; alternate name radB; a frame- shift error in GenBank Accession Number Y00357 changes the C-terminal end of the gene [Escherichia coli]		nuclease	1049
1206	33207450 (2075, 2076)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1026
1207	85817957 (5725, 5726)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1019
1208	65978317 (8299, 8300)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1054
1209	35997388 (8415, 8416)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1029
1210	65500453 (8429, 8430)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1016
1211	78466482 (9839, 9840)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1039
1212	94148304 (10521, 10522)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1045
1213	46782594 (13227, 13228)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1021
1214	19532852 (15337, 15338)	Novel Protein sim. GBank gi 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1034

1215	30519385 (20233) (20234)	Novel Protein sim. GBank g 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1024
1216	52428661 (20867, 20868)	Novel Protein sim. GBank g 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		nuclease	1041
1217	56331119 (3275, 3276)	Novel Protein sim. GBank g 106903 pir B28096 - line-1 protein ORF2 - human		nuclease	1046
1218	37006067 (18843, 18844)	Novel Protein sim. GBank g 106903 pir B28096 - line-1 protein ORF2 - human		nuclease	1029
1219	17948690 (20615, 20616)	Novel Protein sim. GBank g 106903 pir B28096 - line-1 protein ORF2 - human		nuclease	1038
1220	126811096 (21527, 21528)	Novel Protein sim. GBank g 1073852 pir F64055 - anticodon nuclease masking-agent (pruD) homolog - Haemophilus influenzae (strain Rd KN20)		nuclease	1006
1221	85802884 (4845, 4846)	Novel Protein sim. GBank g 111814 pir S21347 - hypothetical protein 3 - rat		nuclease	1054
1222	27956074 (19673, 19674)	Novel Protein sim. GBank g 111816 pir S21348 - probable polyprotein-related protein 4 - rat		nuclease	1004, 1023
1223	38399353 (2097, 2098)	Novel Protein sim. GBank g 112261 pir S16783 - hypothetical protein 2 - rat transposon L1		nuclease	1026
1224	28799359 (8829, 8830)	Novel Protein sim. GBank g 112261 pir S16783 - hypothetical protein 2 - rat transposon L1		nuclease	1022
1225	66437563 (17747, 17748)	Novel Protein sim. GBank g 112261 pir S16783 - hypothetical protein 2 - rat transposon L1		nuclease	1024
1226	5799850 (19179, 19180)	Novel Protein sim. GBank g 112261 pir S16783 - hypothetical protein 2 - rat transposon L1		nuclease	1058

1227	6656375 (20395, 20396)	Novel Protein sim. GBank gil1163120 (U43537) - ORF1; putative ABC excision nuclease repair protein [Streptomyces argillaceus]		nuclease	1022
1228	47657168 (3461, 3462)	Novel Protein sim. GBank gil1169526 sp P44319 END3_HAEIN-ENDONUCLEASE III (DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE)		nuclease	1029
1229	20726633 (10597, 10598)	Novel Protein sim. GBank gil1169543 sp P43728 ERA_HAEIN - GTP-BINDING PROTEIN ERA		nuclease	1006
1230	71729992 (21001, 21002)	Novel Protein sim. GBank gil1169575 sp P43914 EX7S_HAEIN - EXODEOXYRIBONUCLEASE SMALL SUBUNIT (EXONUCLEASE VII SMALL SUBUNIT)		nuclease	1009
1231	27804589 (12917, 12918)	Novel Protein sim. GBank gil1171030 sp P44494 MUTL_HAEIN - DNA MISMATCH REPAIR PROTEIN MUTL		nuclease	1004
1232	9857425 (15301, 15302)	Novel Protein sim. GBank gil119328 sp P20625 END3_ECOLI-ENDONUCLEASE III (DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE)		nuclease	1008
1233	78502974 (1469, 1470)	Novel Protein sim. GBank gil1196431 (M22332) - unknown protein [Homo sapiens]		nuclease	1026
1234	30633991 (10131, 10132)	Novel Protein sim. GBank gil1198431 (M22332) - unknown protein [Homo sapiens]		nuclease	1026
1235	866686833 (5075, 5076)	Novel Protein sim. GBank gil1196307 (L76359) - daunorubicin resistance protein [Streptomyces peucetius]		nuclease	1029
1236	78800554 (19163, 19164)	Novel Protein sim. GBank gil1217671 bbs 173771, human endogenous retrovirus H element-like protein [marmosets, New World monkey, Peptide Partial, 513 aa]		nuclease	1022, 1026
1237	25323521 (8723, 8724)	Novel Protein sim. GBank gil123083 sp P14160 HEXB_STRPN - DNA MISMATCH REPAIR PROTEIN HEXB		nuclease	1008

1238	42725737 (1243, 1244)	Novel Protein sim. GBank gil1237256 (M18247) - gag-pol precursor polyprotein gPr80 [Feline leukemia virus]		nuclease	1050
1239	54996520 (2437, 2438)	Novel Protein sim. GBank gil1262961sp P08547 LIN1_HUMAN - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1001, 1050
1240	65488856 (77, 78)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1054
1241	79926730 (2127, 2128)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1025
1242	348688379 (3945, 3946)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1034
1243	2535644 (4581, 4582)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1009
1244	86680326 (5689, 5690)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1007, 1016
1245	34129168 (15125, 15126)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1026
1246	71450770 (17173, 17174)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1026
1247	65873486 (18235, 18236)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1016
1248	78677286 (18615, 18616)	Novel Protein sim. GBank gil1262961sp P08548 LIN1_NYCCO - LINE-1 REVERSE TRANSCRIPTASE HOMOLOG		nuclease	1026

1249	29673721 (1963, 1964)	Novel Protein sim. GBank gi 130402 sp P11369 POL2_MOUSE - RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]		nuclease	1024
1250	16322200 (17563, 17564)	Novel Protein sim. GBank gi 130584 sp P03360 POL_AVIRE - POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]		nuclease	1034
1251	19631788 (16533, 16534)	Novel Protein sim. GBank gi 131024 sp P17224 PRRD_ECOLI - TYPE I RESTRICTION ENZYME PRRD		nuclease	1004
1252	11295160 (14177, 14178)	Novel Protein sim. GBank gi 131733 sp P25454 RAS51_YEAST - DNA REPAIR PROTEIN RAD51		nuclease	1024
1253	79552113 (5485, 5486)	Novel Protein sim. GBank gi 132256 sp P17894 RECN_BACSU - DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N)		nuclease	1008, 1016, 1039
1254	32120019 (8357, 8358)	Novel Protein sim. GBank gi 133352 sp P09155 RND_ECOLI - RIBONUCLEASE D (RNASE D)		nuclease	1044
1255	94134877 (18917, 18918)	Novel Protein sim. GBank gi 133352 sp P09155 RND_ECOLI - RIBONUCLEASE D (RNASE D)		nuclease	1010
1256	8499144 (9223, 9224)	Novel Protein sim. GBank gi 1335199 emb CAA26919 - (X03145) put. ORF V [Homo sapiens]		nuclease	1034
1257	78305307 (10645, 10646)	Novel Protein sim. GBank gi 1335205 emb CAA36480 - (X52235) ORFII [Homo sapiens]		nuclease	1050
1258	78477054 (13479, 13480)	Novel Protein sim. GBank gi 1335205 emb CAA36480 - (X52235) ORFII [Homo sapiens]		nuclease	1026
1259	6309436 (16545, 16546)	Novel Protein sim. GBank gi 1335205 emb CAA36480 - (X52235) ORFII [Homo sapiens]		nuclease	1058

1260	20440150 (673, 674)	Novel Protein sim. GBank gil1346946 sp P28519 RA14_YEAST - DNA REPAIR PROTEIN RAD14		nuclease	1024
1261	11594285 (9187, 9188)	Novel Protein sim. GBank gil1346946 sp P28519 RA14_YEAST - DNA REPAIR PROTEIN RAD14		nuclease	1024
1262	94135973 (13473, 13474)	Novel Protein sim. GBank gil135078 sp P09959 SWI6_YEAST - REGULATORY PROTEIN SWI6 (CELL-CYCLE BOX FACTOR, CHAIN SWI6) (TRANS-ACTING ACTIVATOR OF HO ENDONUCLEASE GENE) (MBF SUBUNIT P90)		nuclease	1003
1263	36023923 (13553, 13554)	Novel Protein sim. GBank gil1350816 sp P4820 RNP_A_STRCO - RIBONUCLEASE P PROTEIN COMPONENT (PROTEIN C5) (RNASE P)		nuclease	1053
1264	95350945 (22131, 22132)	Novel Protein sim. GBank gil1352824 sp P3085 VACB_SHIFL - VACB PROTEIN		nuclease	1010
1265	20608222 (603, 604)	Novel Protein sim. GBank gil140751 sp P25537 CAFA_ECOLI - CYTOPLASMIC AXIAL FILAMENT PROTEIN		nuclease	1004
1266	27980924 (16769, 16770)	Novel Protein sim. GBank gil1533037 (U68235) - RecN [Bacillus subtilis]		nuclease	1006
1267	25241756 (17741, 17742)	Novel Protein sim. GBank gil1533037 (U68235) - RecN [Bacillus subtilis]		nuclease	1008
1268	66729956 (15701, 15702)	Novel Protein sim. GBank gil1666140 emb CAB03375 - (Z81451) me [Mycobacterium tuberculosis]		nuclease	1010

1269	87587610 (14483, 14484)	Novel Protein sim. GBank gi 170642 1spl P39875 EXO1_YEAST - EXONUCLEASE I (EXO I) (DHs1 PROTEIN)	nuclease	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1026, 1027, 1029, 1031, 1034, 1039, 1044
1270	13526841 (5541, 5542)	Novel Protein sim. GBank gi 170906 1spl P32829 MR11_YEAST - MRE11 PROTEIN	nuclease	1001
1271	29983095 (6709, 6710)	Novel Protein sim. GBank gi 171261 (M37696) - damage-repair protein [Saccharomyces cerevisiae]	nuclease	1006, 1016, 1024
1272	35893235 (19351, 19352)	Novel Protein sim. GBank gi 1718063 1spl P52087 UVRA_PSELE - PUTATIVE EXCINUCLEASE_ABC SUBUNIT A	nuclease	1053
1273	78905409 (14249, 14250)	Novel Protein sim. GBank gi 172002 (M841170) - homologue of bacterial MutS protein [Saccharomyces cerevisiae]	nuclease	1009, 1039
1274	20465183 (3931, 3932)	Novel Protein sim. GBank gi 172203 (M29688) - DNA mismatch repair protein [Saccharomyces cerevisiae]	nuclease	1010
1275	25131847 (19309, 19310)	Novel Protein sim. GBank gi 172385 (M22222) - REV1 protein [Saccharomyces cerevisiae]	nuclease	1008
1276	326955308 (14739, 14740)	Novel Protein sim. GBank gi 189479 1emb CAA62819 - (X91617) 5'-3' exonuclease [Mus musculus]	nuclease	1000
1277	86682025 (5171, 5172)	Novel Protein sim. GBank gi 1916229 (U83303) - line-1 reverse transcriptase [Homo sapiens]	nuclease	1007
1278	38066565 (8641, 8642)	Novel Protein sim. GBank gi 198646 (M29324) - ORF2; 5' end undetermined [Mus musculus]	nuclease	1022
1279	16554637 (17719, 17720)	Novel Protein sim. GBank gi 198646 (M29324) - ORF2; 5' end undetermined [Mus musculus]	nuclease	1023

1280 5884	29442333 (5883, putative p150 [Homo sapiens])	Novel Protein sim. GBank gil2072948 (U93563) - putative p150 [Homo sapiens]		nuclease	1030
1281 20008030 (9461, 9462)	Novel Protein sim. GBank gil2072948 (U93563) - putative p150 [Homo sapiens]			nuclease	1022
1282 366622017 (21313, 21314)	Novel Protein sim. GBank gil2072948 (U93563) - putative p150 [Homo sapiens]			nuclease	1034
1283 355345253 (1195, 1196)	Novel Protein sim. GBank gil2072961 (U93568) - putative p150 [Homo sapiens]			nuclease	1026
1284 30401845 (17877, 17878)	Novel Protein sim. GBank gil2072961 (U93568) - putative p150 [Homo sapiens]			nuclease	1026
1285 34624846 (20781, 20782)	Novel Protein sim. GBank gil2072961 (U93568) - putative p150 [Homo sapiens]			nuclease	1029
1286 6631952 (18375, 18376)	Novel Protein sim. GBank gil2072964 (U93569) - putative p150 [Homo sapiens]			nuclease	1022
1287 79116463 (2553, 2584)	Novel Protein sim. GBank gil2072967 (U93570) - putative p150 [Homo sapiens]			nuclease	1013
1288 34073815 (6127, 6128)	Novel Protein sim. GBank gil2072967 (U93570) - putative p150 [Homo sapiens]			nuclease	1012
1289 33184219 (14903, 14904)	Novel Protein sim. GBank gil2072967 (U93570) - putative p150 [Homo sapiens]			nuclease	1028
1290 78252004 (10309, 10310)	Novel Protein sim. GBank gil2072972 (U93572) - putative p150 [Homo sapiens]			nuclease	1026
1291 32419490 (19593, 19594)	Novel Protein sim. GBank gil2072972 (U93572) - putative p150 [Homo sapiens]			nuclease	1012
1292 8367492 (7427, 7428)	Novel Protein sim. GBank gil2193870 dbj BAA20419 - (D84391) reverse transcriptase [Mus musculus]			nuclease	1010
1293 65680323 (12035, 12036)	Novel Protein sim. GBank gil2193870 dbj BAA20419 - (D84391) reverse transcriptase [Mus musculus]			nuclease	1016
1294 35049576 (22399, 22400)	Novel Protein sim. GBank gil2193870 dbj BAA20419 - (D84391) reverse transcriptase [Mus musculus]			nuclease	1030, 1038
1295 25259684 (12735, 12736)	Novel Protein sim. GBank gil2197085 (AF003535) - ORF2-like protein [Homo sapiens]			nuclease	1026

1296	91232168 (739, 740)	Novel Protein sim. GBank gil225047 prfl 1207289A - reverse transcriptase related protein [Homo sapiens]		nuclease	1010
1297	78479498 (14081, 14082)	Novel Protein sim. GBank gil2343271 (AF014793) - pol polyprotein [walleye epidermal hyperplasia virus type 2]		nuclease	1026
1298	79833737 (3729, 3730)	Novel Protein sim. GBank gil2497995 spI Q56215 MUTS_THEAQ - DNA MISMATCH REPAIR PROTEIN MUTS		nuclease	1008
1299	79841635 (14203, 14204)	Novel Protein sim. GBank gil249881 spI Q51424 RUV_C_PSEAE - CROSSOVER JUNCTION ENDODEOXYRIBONUCLEASE RUV_C (HOLLIDAY JUNCTION NUCLEASE RUV_C) (HOLLIDAY JUNCTION RESOLVASE RUV_C)		nuclease	1039, 1044
1300	78981831 (19889, 19890)	Novel Protein sim. GBank gil26701 spI P30620 PSO2_YEAST - DNA CROSS-LINK REPAIR PROTEIN PSO2/SNM1		nuclease	1017, 1044
1301	8002657 (14479, 14480)	Novel Protein sim. GBank gil2791504 emb CAA16041 - (AL021246) hypothetical protein Rv2464c [Mycobacterium tuberculosis]		nuclease	1006, 1017
1302	78896818 (13677, 13678)	Novel Protein sim. GBank gil2815315 emb CAA16438 - (AL021529) hypothetical protein SC10A5.06 [Streptomyces coelicolor]		nuclease	1008, 1034, 1054
1303	21414899 (51, 52)	Novel Protein sim. GBank gil2979589 (AC004416) - WUGSC:H_RGO13N12.gw.1335199. a gene product [Homo sapiens]		nuclease	1034
1304	78091048 (2787, 2788)	Novel Protein sim. GBank gil29816311 dbj BAA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1037
1305	57310383 (8215, 8216)	Novel Protein sim. GBank gil2981631 dbj BAA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1024

1306	65874248 (10289, 10290)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1016
1307	80188936 (10779, 10780)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1058
1308	57459107 (11667, 11668)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1054
1309	17897645 (12967, 12968)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1044
1310	33194820 (15689, 15690)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1026
1311	47662433 (20257, 20258)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1037
1312	43335214 (20919, 20920)	Novel Protein sim. GBank gi 2981631 dbj BA25253.1 - (AB012223) ORF2 [Canis familiaris]		nuclease	1037, 1050
1313	13085963 (3455, 3456)	Novel Protein sim. GBank gi 30252 18 sp Q47685 YKFG_ECOLI - HYPOTHETICAL 18.1 KD PROTEIN IN PROAPERR INTERGENIC REGION		nuclease	1022
1314	79571269 (1593, 1594)	Novel Protein sim. GBank gi 3089615 (AF086791) - ribonuclease H [Zymomonas mobilis]		nuclease	1017
1315	19513725 (5137, 5138)	Novel Protein sim. GBank gi 3122992 sp O06150 UVRB_MYCTU - EXCINUCLEASE ABC SUBUNIT B		nuclease	1023
1316	86685818 (19445, 19446)	Novel Protein sim. GBank gi 3122992 sp O06150 UVRB_MYCTU - EXCINUCLEASE ABC SUBUNIT B		nuclease	1024, 1029

1317	13514245 (17366) 17366)	Novel Protein sim. GBank gi 3150225 emb CAA19197 - (AL023635) formamidopyrimidine-dna glycosylase [Mycobacterium leprae]		nuclease	1016, 1024
1318	20482144 (18087, 18088)	Novel Protein sim. GBank gi 3219304 dbj BAA238847 - (AB009461) MUS38 [Neurospora crassa]		nuclease	1010
1319	94327229 (3871, 3872)	Novel Protein sim. GBank gi 325455 (M74509) - [Human endogenous retrovirus type C oncovirus sequence.], gene product [Homo sapiens]		nuclease	1010
1320	80502370 (19053, 19054)	Novel Protein sim. GBank gi 3261599 emb CAB00917 - (277137) hypothetical protein Rv1277 [Mycobacterium tuberculosis]		nuclease	1012, 1026, 1031, 1044
1321	34405865 (15397, 15398)	Novel Protein sim. GBank gi 3449248 emb CAA20392 - (AL031317) putative phospho-sugar mutase [Streptomyces coelicolor]		nuclease	1030
1322	80028104 (11819, 11820)	Novel Protein sim. GBank gi 3581916 emb CAA20855 - (AL031545) mutS family DNA mismatch repair protein [Schizosaccharomyces pombe]		nuclease	1006, 1024
1323	50665154 (9677, 9678)	Novel Protein sim. GBank gi 3569347 (AF081114) - ORF2 [Mus musculus domesticus]		nuclease	1042
1324	55181838 (49, 50)	Novel Protein sim. GBank gi 3860767 emb CAA14668 - (AJ235270) EXCINUCLEASE ABC SUBUNIT B (uvrB) [Rickettsia prowazekii]		nuclease	1019
1325	65902625 (11075, 11076)	Novel Protein sim. GBank gi 386018 emb CAA14918 - (AJ235271) RIBONUCLEASE D (rnd) [Rickettsia prowazekii]		nuclease	1023

1326	11076703 (10277, 10278)	Novel Protein sim. G-Bank g 3861120 emb CAA15020 - (AJ235272) EXCINUCLEASE ABC SUBUNIT C (uvrC) [Rickettsia prowazekii]		nuclease	1010
1327	66258962 (13847, 13848)	Novel Protein sim. G-Bank g 3915204 spi O34863 UVRA_BACSU - EXCINUCLEASE ABC SUBUNIT A		nuclease	1010
1328	9892207 (19047, 19048)	Novel Protein sim. G-Bank g 3915218 spi O31151 UVRA_ZYMMO - EXCINUCLEASE ABC SUBUNIT A		nuclease	1034
1329	80504703 (4877, 4878)	Novel Protein sim. G-Bank g 400965 spi P31337 RADC_STAAU - DNA REPAIR PROTEIN RADC HOMOLOG (25 KD PROTEIN)		nuclease	1012
1330	39575286 (7155, 7156)	Novel Protein sim. G-Bank g 4020 emb CAA28789 - (X05443) beginning of CDS on complementary strand with homology to recF of E.coli (PNAS 81 : 4622-4626) [Saccharomyces cerevisiae]		nuclease	1031
1331	78741313 (15787, 15788)	Novel Protein sim. G-Bank g 417587 spi P32849 RAD5_YEAST - DNA REPAIR PROTEIN RAD5		nuclease	1008, 1026
1332	10869337 (20071, 20072)	Novel Protein sim. G-Bank g 423680 pir S33477 - hypothetical protein 1 - rat (fragment)		nuclease	1024
1333	78738979 (2837, 2838)	Novel Protein sim. G-Bank g 439877 L27428 - reverse transcriptase [Homo sapiens]		nuclease	1044
1334	80162752 (5669, 5670)	Novel Protein sim. G-Bank g 4539097 emb CAE39815.1 - (AL049491) putative exodeoxyribonuclease small subunit [Mycobacterium leprae]		nuclease	1008, 1058
1335	78778092 (12455, 12456)	Novel Protein sim. G-Bank g 460627 U07187 - Mlh1p [Saccharomyces cerevisiae]		nuclease	1009
1336	787786855 (18821, 18822)	Novel Protein sim. G-Bank g 465488 spi P33301 XRS2_YEAST - DNA REPAIR PROTEIN XRS2		nuclease	1022, 1026

1337	25258159 (14603, 14604)	Novel Protein sim. GBank gil482970 pir B388888 - COI Inton 9 protein - Podospora anserina mitochondrion (SGC3)		nuclease	1026
1338	29501308 (22155, 22156)	Novel Protein sim. GBank gil5031889 refNP_005570.1 pLRE1 - LINE retrotransposable element 1		nuclease	1016
1339	30793972 (2311, 2312)	Novel Protein sim. GBank gil509838 (M165550) - pol polyprotein [Baboon endogenous virus]		nuclease	1026
1340	87587598 (2265, 2266)	Novel Protein sim. GBank gil545817 bbs1145945 - (S68545) Dhs1 [Saccharomyces cerevisiae, Peptide, 483 aa] [Saccharomyces cerevisiae]		nuclease	1003, 1004, 1008, 1009, 1010, 1022, 1026, 1039, 1044
1341	30790837 (8061, 8062)	Novel Protein sim. GBank gil558414 emb CAA86260) - (Z38114) len: 750, CAI: 0.14, incomplete ORF, PIF YEAST P0727 1 MITOCHONDRIAL DNA REPAIR AND RECOMBINATION PROTEIN PIF1 PRECURSOR [Saccharomyces cerevisiae]		nuclease	1009
1342	25144845 (6903, 6904)	Novel Protein sim. GBank gil5690058 emb CAB51943.1 - (AJ243915) excinuclease, subunit A [Staphylococcus xylosus]		nuclease	1026
1343	56120421 (15857, 15888)	Novel Protein sim. GBank gil599989 emb CAA86864) - (Z46833) unknown [Saccharomyces cerevisiae]		nuclease	1017
1344	11394335 (10617, 10618)	Novel Protein sim. GBank gil731870 sp P40469 MT18_YEAST - DNA REPAIR/TRANSCRIPTION PROTEIN MET18/MMS19		nuclease	1034
1345	95292719 (12123, 12124)	Novel Protein sim. GBank gil79839 pir S03812 - lvrB protein - Micrococcus luteus		nuclease	1006, 1022, 1024, 1026, 1034
1346	56512843 (9535, 9536)	Novel Protein sim. GBank gil804608 (M13100) - unknown protein [Rattus norvegicus]		nuclease	1019
1347	60196735 (3631, 3632)	Novel Protein sim. GBank gil804611 (M13100) - unknown protein [Rattus norvegicus]		nuclease	1026, 1034
1348	71754301 (845, 846)	Novel Protein sim. GBank gil807677 (M13101) - unknown protein [Rattus norvegicus]		nuclease	1003

1349	30509050 (13341, 13342)	Novel Protein sim. GBank gil807677 (M13101) - unknown protein [Rattus norvegicus]		nuclease	1009
1350	32291115 (17127, 17128)	Novel Protein sim. GBank gil807677 (M13101) - unknown protein [Rattus norvegicus]		nuclease	1000
1351	55004978 (343, 344)	Novel Protein sim. GBank gil87765 pir JU0033 - hypothetical L1 protein (third intron of gene TS) - human		nuclease	1019
1352	33700472 (541, 542)	Novel Protein sim. GBank gil87765 pir JU0033 - hypothetical L1 protein (third intron of gene TS) - human		nuclease	1026
1353	86476763 (2263, 2264)	Novel Protein sim. GBank gil87765 pir JU0033 - hypothetical L1 protein (third intron of gene TS) - human		nuclease	1029
1354	30536167 (14779, 14780)	Novel Protein sim. GBank gil87765 pir JU0033 - hypothetical L1 protein (third intron of gene TS) - human		nuclease	1026
1355	80029899 (22647, 22648)	Novel Protein sim. GBank gil87765 pir JU0033 - hypothetical L1 protein (third intron of gene TS) - human		nuclease	1026
1356	78389591 (785, 786)	Novel Protein sim. GBank gil92728 pir PH0217 - reverse transcriptase-like protein - rat (fragment)		nuclease	1008
1357	25258621 (9479, 9480)	Novel Protein sim. GBank gil92728 pir PH0217 - reverse transcriptase-like protein - rat (fragment)		nuclease	1008
1358	35104966 (9485, 9486)	Novel Protein sim. GBank gil92728 pir PH0217 - reverse transcriptase-like protein - rat (fragment)		nuclease	1029
1359	66470568 (19701, 19702)	Novel Protein sim. GBank gil92728 pir PH0217 - reverse transcriptase-like protein - rat (fragment)		nuclease	1009
1360	54530714 (22837, 22838)			nuclease	1041
1361	79830303 (10533, 10534)	Novel Protein sim. GBank gil117422 sp P10040 CRB_DROME - CRUMBS PROTEIN PRECURSOR (95F)	Contains protein domain (PF00008) - EGF-like domain	oncogene	1008, 1009, 1038, 1039, 1054

1362	15028498 (9433, 9434)	Novel Protein sim. GBank gil 352514 sp P07207 NOTC_DROME - NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR	Contains protein domain (PF00008) - EGF-like domain	oncogene	1054
1363	35056380 (19885, 19886)	Novel Protein sim. GBank gil 3449286 dbj BA-A32458 - (AB011527) MEGF1 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	oncogene	1000
1364	9646220 (14405, 14406)	Novel Protein sim. GBank gil 552087 (M33753) - crumbs protein [Drosophila melanogaster]	Contains protein domain (PF00008) - EGF-like domain	oncogene	1039
1365	80098977 (20105, 20102)	Novel Protein sim. GBank gil 584850 sp P38041 BOB1 YEAST - BOB1 PROTEIN (BEM1-BINDING PROTEIN)	Contains protein domain (PF00018) - SH3 domain	oncogene	1006, 1010, 1012, 1016, 1017, 1022, 1024, 1025, 1026, 1039
1366	79758899 (5803, 5804)	Novel Protein sim. GBank gil 477495 pir A49128 - cell-fate determining gene Notch2 protein - rat	Contains protein domain (PF00066) - Notch (DSL) domain	oncogene	1001, 1008
1367	78912460 (4751, 4752)	Novel Protein sim. GBank gil 353108 sp P48566 YN33_YEAST - HYPOTHETICAL 73.0 KD PROTEIN IN CLA4-MID1 INTERGENIC REGION	Contains protein domain (PF00566) - TBC domain	oncogene	1003, 1008, 1009, 1017, 1026, 1039
1368	20634766 (11201, 11202)	Novel Protein sim. GBank gil 391463 sp P74932 RGSE_MOUSE - REGULATOR OF G-PROTEIN SIGNALING 14 (RGS14) (RAP1/RAP2 INTERACTING PROTEIN)	Contains protein domain (PF00615) - Regulator of G protein signaling domain	oncogene	1022
1369	65698315 (21763, 21764)	Novel Protein sim. GBank gil 076712 (U34932) - Fos-related antigen (Rattus norvegicus)		oncogene	1038
1370	20708062 (3295, 3296)	Novel Protein sim. GBank gil 106322 pir B34087 - hypothetical protein (L1H 3' region) - human		oncogene	1006
1371	28385329 (19823, 19824)	Novel Protein sim. GBank gil 119134 sp P06805 EF11_RHRA - ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)		oncogene	1008

1372	79652286 (8741, 8742)	Novel Protein sim. GBank gil135914 sp P7408 TLM_MOUSE - TLM PROTEIN (TLM ONCOGENE)		oncogene	1003, 1008, 1009, 1017, 1022, 1023, 1026, 1039, 1044
1373	20289691 (18145, 18146)	Novel Protein sim. GBank gil3702962 (AF079484) - rac GTP binding protein Arac7 [Arabidopsis thaliana]		oncogene	1022
1374	80570966 (12635, 12636)	Novel Protein sim. GBank gil4826962 ref NP_005043.1 pRAC3 - ras-related C3 botulinum toxin substrate 3		oncogene	1058
1375	27970162 (955, 956)	Novel Protein sim. GBank gil5081604 gb AAD39462.1 AF13512 - (AF-135-125) nuclear factor kappa B subunit p100 [Mus musculus]		oncogene	1034
1376	11104008 (3479, 3480)	Novel Protein sim. GBank gil5081604 gb AAD39462.1 AF13512 - (AF-135-125) nuclear factor kappa B subunit p100 [Mus musculus]		oncogene	1022
1377	20297291 (1465, 1466)	Novel Protein sim. GBank gil1346011 sp Q03073 FIXN_BRAJA - CYTOCHROME C OXIDASE POLYPEPTIDE I HOMOLOG, BACTERIUM (CYTOCHROME CBB3 SUBUNIT 1) (HEME B/COPPER CYTOCHROME C OXIDASE SUBUNIT)	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1034
1378	20292836 (13933, 13934)	Novel Protein sim. GBank gil1346012 sp Q05572 FIXN_RHIME - CYTOCHROME C OXIDASE POLYPEPTIDE I HOMOLOG, BACTERIUM (CYTOCHROME CBB3 SUBUNIT 1) (HEME B/COPPER CYTOCHROME C OXIDASE SUBUNIT)	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1034
1379	95293310 (16757, 16758)	Novel Protein sim. GBank gil1352143 sp P98001 ICOX1_SACDO - CYTOCHROME C OXIDASE POLYPEPTIDE I	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1004, 1006, 1010, 1022, 1024, 1034

1380	78362231 (2529, 2530)	Novel Protein sim. GBank gil4001773 (AF037980) - cytochrome C oxidase subunit I; COXI [Chaitinella antiqua]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1008
1381	25311332 (17067, 17068)	Novel Protein sim. GBank gil4154173 (U96639) - cytochrome c oxidase subunit I [Canis familiaris]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1009
1382	94121065 (20473, 20474)	Novel Protein sim. GBank gil4239837[dbj BAA74723] - (AB016734) Cytochrome c oxidase subunit I [Tarsius bancanus]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1016, 1044
1383	80504718 (7321, 7322)	Novel Protein sim. GBank gil4539224[emb CAB39882.1] - (AL049497) putative cytochrome c oxidase subunit I [Streptomyces coelicolor]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1012
1384	25244108 (21953, 21954)	Novel Protein sim. GBank gil4539224[emb CAB39882.1] - (AL049497) putative cytochrome c oxidase subunit I [Streptomyces coelicolor]	Contains protein domain (PF00115) - Cytochrome C and Quinol oxidase polypeptide I	oxidase	1008
1385	47658259 (19609, 19610)	Novel Protein sim. GBank gil4539225[emb CAB39883.1] - (AL049497) putative cytochrome c oxidase subunit II [Streptomyces coelicolor]	Contains protein domain (PF00116) - Cytochrome C oxidase subunit II	oxidase	1029, 1053
1386	87461740 (18611, 18612)	Novel Protein sim. GBank gil3212610[pdb 1SOX A - Chain A, Sulfite Oxidase From Chicken Liver	Contains protein domain (PF00174) - Oxidoreductase molybdopterin binding domain	oxidase	1025
1387	65645753 (15749, 15750)	Novel Protein sim. GBank gil140821[sp P10343 Y142_PSEAY - HYPOTHETICAL 42.6 KD PROTEIN IN ISOAMYLASE 3 REGION	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	oxidase	1041
1388	13085734 (20559, 20560)	Novel Protein sim. GBank gil4007704[emb CAA22389] - (AL034446) pyruvate dehydrogenase [Streptomyces coelicolor]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	oxidase	1027

1389	78362389 (2455, 2456)	Novel Protein sim. GBank g 1495802 emb CAA65268I - (X96405) 13-lipoxygenase [Solanum tuberosum]	Contains protein domain (PF00305) - Lipoxygenase oxidase	1003
1390	86688955 (16579, 16580)	Novel Protein sim. GBank g 21822671gb AAB60715.1 - (L37559) lipoxygenase [Hordeum vulgare]	Contains protein domain (PF00305) - Lipoxygenase oxidase	1007
1391	23330421 (9305, 9306)	Novel Protein sim. GBank g 2505825 sp P29114 LOX1_HORVU - LIPOXYGENASE 1	Contains protein domain (PF00305) - Lipoxygenase oxidase	1053
1392	27354903 (19127, 19128)	Novel Protein sim. GBank g 1175947 sp P43561 YFE1_YEAST - PUTATIVE MULTICOPPER OXIDASE YFD041W PRECURSOR	Contains protein domain (PF00394) - Multicopper oxidase	1010, 1022, 1026
1393	79462975 (21327, 21328)	Novel Protein sim. GBank g 1175947 sp P43561 YFE1_YEAST - PUTATIVE MULTICOPPER OXIDASE YFD041W PRECURSOR	Contains protein domain (PF00394) - Multicopper oxidase	1010, 1022, 1026
1394	78378111 (14621, 14622)	Novel Protein sim. GBank g 1317201 sp P25996 PYRD_BACSU - ASPARTATE OXIDASE (QUINOLINATE SYNTHETASE B)	Contains protein domain (PF01180) - Dihydroorotate domain	1031
1395	23207489 (7723, 7724)	Novel Protein sim. GBank g 1352167 sp P12075 COXB_RAT - CYTOCHROME C OXIDASE POLYPEPTIDE-VB PRECURSOR (VIa')	Contains protein domain (PF01215) - Cytochrome c oxidase subunit Vb oxidase	1008
1396	79247420 (8569, 8570)	Novel Protein sim. GBank g 1170220 sp P43898 HEM6_PSEAE - COPROPORPHYRINOGEN III OXIDASE, AEROBIC (COPROPORPHYRINOGENASE) (COPROGEN OXIDASE)	Contains protein domain (PF01215) - Cytochrome c oxidase subunit Vb oxidase	1039
1397	80036992 (22011, 22012)	Novel Protein sim. GBank g 1175938 sp P45106 YFCB_HAEIN - HYPOTHETICAL ADENINE-SPECIFIC METHYLASE_H1201	Contains protein domain (PF01215) - Cytochrome c oxidase subunit Vb oxidase	1010

1398	70287770 (499, 500)	Novel Protein sim. GBank g 128545 sp P24474 NIRS_PSEAE - NITRITE REDUCTASE PRECURSOR (CYTOCHROME CD1) (CYTOCHROME OXIDASE)		oxidase	1009
1399	20439488 (7387, 7388)	Novel Protein sim. GBank g 131719 sp P28294 PYRD_AGRAE - DIHYDROORotate DEHYDROGENASE (DIHYDROORotate OXIDASE) (DHODEHASE)		oxidase	1024
1400	78459744 (18893, 18894)	Novel Protein sim. GBank g 1575339 (U60417) - PapM [Streptomyces pristinaespiralis]		oxidase	1024, 1026
1401	66002714 (2833, 2834)	Novel Protein sim. GBank g 17060541 sp Q10375 COX2_MYCTU - PUTATIVE CYTOCHROME C OXIDASE POLYPEPTIDE II PRECURSOR (CYTOCHROME AA3 SUBUNIT 2)		oxidase	1010
1402	- 20457608 (21797, 21798)	Novel Protein sim. GBank g 1707879 sp P54261 GCST_SYNY3 - PROBABLE AMINOMETHYLTRANSFERASE (GLYCINE CLEAVAGE SYSTEM T PROTEIN)		oxidase	1010
1403	9504926 (2047, 2048)	Novel Protein sim. GBank g 172217 (M27515) - acyl-coenzyme A oxidase [Saccharomyces cerevisiae]		oxidase	1003, 1008, 1009, 1039, 1044
1404	20724144 (12855, 12856)	Novel Protein sim. GBank g 2126013 sp S53665 - proline dehydrogenase (EC 1.5.99.8) / 1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12) - Escherichia coli		oxidase	1006
1405	94993394 (7293, 7294)	Novel Protein sim. GBank g 2135870 sp I38079 - OXA1 homolog - human		oxidase	1003, 1004, 1006, 1022, 1024, 1026, 1027, 1030, 1034, 1053
1406	27803758 (13823, 13824)	Novel Protein sim. GBank g 2208964 emb CAAA71555 - (Y10528) cyanide insensitive terminal oxidase [Pseudomonas aeruginosa]		oxidase	1010

1407	858 8174 (7549, 7550)	Novel Protein sim. GBank gi 2208965 emb CAA71556 - (Y10528) cyanide insensitive terminal oxidase [Pseudomonas aeruginosa]		oxidase	1019
1408	21426551 (8103, 8104)	Novel Protein sim. GBank gi 2493813 sp P71756 HEMIN_MYCTU - PROBABLE OXYGEN-INDEPENDENT COPROPORPHYRINogen III OXIDASE (COPROPORPHYRINogenase) (COPROGEN OXIDASE)		oxidase	1022
1409	79262236 (22067, 22068)	Novel Protein sim. GBank gi 2498208 sp Q63448 CAOQ_RAT - PRISTANOYL-COA OXIDASE		oxidase	1003, 1008
1410	9282157 (1115, 1116)	Novel Protein sim. GBank gi 2498608 sp Q51363 NADB_PSEAE -L- ASPARTATE OXIDASE (QUINOLINATE SYNTETASE B)		oxidase	1008
1411	65685771 (4455, 4456)	Novel Protein sim. GBank gi 2498933 sp Q46337 SOXA_CORSP - SARCOSINE OXIDASE ALPHA SUBUNIT		oxidase	1024
1412	20446820 (16391, 16392)	Novel Protein sim. GBank gi 2498935 sp Q46338 SOXG_CORSP - SARCOSINE OXIDASE GAMMA SUBUNIT		oxidase	1024
1413	79833670 (22977, 22978)	Novel Protein sim. GBank gi 2506687 sp P33225 [TORA_ECOLI - TRIMETHYLAMINE-N-OXIDE REDUCTASE PRECURSOR (TMAO REDUCTASE) (TRIMETHYLAMINE OXIDASE)]		oxidase	1008
1414	16282886 (8749, 8750)	Novel Protein sim. GBank gi 2588164 dbj BAA23227 - (D88452) aldehyde oxidase-2 [Zea mays]		oxidase	1024
1415	80237288 (8191, 8192)	Novel Protein sim. GBank gi 2605606 dbj BAA23329 - (D50559) RANP-1 [Rattus norvegicus]		oxidase	1004

1416	79408531 (11583, 11584)	Novel Protein sim. GBank gi 2624303 emb CAA15576I - (AL008967) hypothetical protein Rv2781c [Mycobacterium tuberculosis] - germin-like protein 1 [<i>Oryza sativa</i>]		oxidase	1016
1417	28460575 (18515, 18516)	Novel Protein sim. GBank gi 2655285 (AF032971) predicted protein [<i>Arabidopsis thaliana</i>]		oxidase	1024
1418	7660740 (14287, 14288)	Novel Protein sim. GBank gi 2827552 emb CAA16560I - (AL021635) predicted protein [<i>Arabidopsis thaliana</i>]		oxidase	1034
1419	17947247 (13939, 13940)	Novel Protein sim. GBank gi 2828219 sp P38993 FET3_YEAST - IRON TRANSPORT MULTICOPPER OXIDASE PRECURSOR		oxidase	1039
1420	78733845 (22721, 22722)	Novel Protein sim. GBank gi 2828219 sp P38993 FET3_YEAST - IRON TRANSPORT MULTICOPPER OXIDASE PRECURSOR		oxidase	1026, 1034
1421	79789320 (8409, 8410)	Novel Protein sim. GBank gi 3023547 sp P56392 COXK_MOUSE - CYTOCHROME C OXIDASE POLYPEPTIDE VIIA HEART PRECURSOR (COX VIIA-M)		oxidase	1006, 1017
1422	94233606 (8681, 8682)	Novel Protein sim. GBank gi 3294501 (U64857) - similar to the DPT1/Kunitz family of inhibitors; most similar to tissue factor pathway inhibitor precursor [<i>Caenorhabditis elegans</i>]		oxidase	1010
1423	9884328 (10527, 10528)	Novel Protein sim. GBank gi 3378433 (AF079317) - flavoprotein subunit p-cresol methylhydroxylase [Sphingomonas aromaticivorans]		oxidase	1039
1424	37800703 (9609, 9610)	Novel Protein sim. GBank gi 3550333 (AF084032) - quinol oxidase subunit II QoxB [Rhodobacter sphaeroides]		oxidase	1012
1425	78758894 (4791, 4792)	Novel Protein sim. GBank gi 3914407 sp O55230 PPOX_MYCTU - PROTOPORPHYRINOGEN OXIDASE (PPO)		oxidase	1008

1426	79574267 (5849, 5850)	Novel Protein sim. GBank gi 39153971 sp Q55629 Y782_SYNY3 - HYPOTHETICAL 51.4 KD PROTEIN SLR0782		oxidase	1031
1427	66395745 (6105, 6106)	Novel Protein sim. GBank gi 417339 sp P32382 NADH_OXIDASE		oxidase	1008
1428	78890741 (18995, 18996)	Novel Protein sim. GBank gi 425159 (M24389) - cytochrome c oxidase subunit V [Neurospora crassa]		oxidase	1039
1429	82474822 (18731, 18732)	Novel Protein sim. GBank gi 4323583 gb AAD16434 - (AF101076) maturose-related protein [Pseudomonas putida]		oxidase	1001, 1025
1430	23329863 (13005, 13006)	Novel Protein sim. GBank gi 4514628 dbj BAA75482.1 - (ABD16894) bd-type quinol oxidase subunit I [Bacillus stearothermophilus]		oxidase	1026
1431	71091932 (4861, 4862)	Novel Protein sim. GBank gi 4539225 emb CAB39883.1 - (AL049497) putative cytochrome c oxidase subunit II [Streptomyces coelicolor]		oxidase	1053
1432	88089761 (14633, 14634)	Novel Protein sim. GBank gi 4581879 gb AAD24776.1 AF12027 - (AF120279) proline dehydrogenase; PRODH [Mus musculus]		oxidase	1053
1433	71827009 (13517, 13518)	Novel Protein sim. GBank gi 465011 sp P33282 URIC_EMENI - URICASE (URATE_OXIDASE)		oxidase	1009
1434	30790524 (5753, 5754)	Novel Protein sim. GBank gi 473729 dbj BAA01743 - (D10951) cytochrome c oxidase subunit Vb [Rattus rattus]		oxidase	1026
1435	13088832 (16189, 16190)	Novel Protein sim. GBank gi 497637 (J03939) - cytochrome oxidase d subunit I [Escherichia coli]		oxidase	1022, 1023

1436	27929985 (21649, 21650)	Novel Protein sim. GBank gil544323 sp P36365 FMO1_RAT - DIMETHYLANILINE MONOOXYGENASE (N-OXIDE FORMING) 1 (HEPATIC FLAVIN-CONTAINING MONOOXYGENASE 1) (FMO 1) (DIMETHYLANILINE OXIDASE 1)		oxidase	1008
1437	30783191 (4903, 4904)	Novel Protein sim. GBank gil585656 sp P38075 PDX3_YEAST - PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)		oxidase	1009
1438	27981877 (2791, 2792)	Novel Protein sim. GBank gil731440 sp P400121 PPOX_YEAST - PROTOPORPHYRINogen OXIDASE (PPO)		oxidase	1006
1439	23299174 (5225, 5226)			oxidase	1044
1440	37036378 (20769, 20770)			oxidase	1012
1441	10885177 (12249, 12250)	Novel Protein sim. GBank gil3023935 sp P77266 HCAC_ECOLI - DIGOXIGENIN FERREDOXIN SUBUNIT	Contains protein domain (PF00355) - Rieske [2Fe-2S] domain	oxygenase	1027
1442	88094187 (2247, 2248)	Novel Protein sim. GBank gil4503011 ref NP_001299.1 pCPN1 - carboxy peptidase N, polypeptide 1, 50kD precursor	Contains protein domain (PF00246) - Zinc carboxy peptidase	peptidase	1006, 1022
1443	80028158 (20893, 20894)	Novel Protein sim. GBank gil465787 sp P34422 YL31_CAEEL - HYPOTHETICAL 86.0 KD PROTEIN F44B9.1 IN CHROMOSOME III	Contains protein domain (PF00326) - Prolyl oligopeptidase family	peptidase	1006, 1025
1444	65681075 (11219, 11220)	Novel Protein sim. GBank gil4186118 emb CAAA10649 - (AJ132364) type 4 prephilin peptidase [Pseudomonas stutzeri]	Contains protein domain (PF01478) - Type III leader peptidase family	peptidase	1041
1445	23304605 (7605, 7606)	Novel Protein sim. GBank gil113739 sp P19994 AMPM_BACSU - METHIONINE AMINOPEPTIDASE (MAP) (PEPTIDASE M)		peptidase	1008

1446 3812)	79606095 (3811, Novel Protein sim. GBank g 168448 sp Q5813 AMP1_STRLI - XAA-PRO AMINOPEPTIDASE I (X-PRO AMINOPEPTIDASE I) (AMINOPEPTIDASE P I) (APP) (PEPP I) (AMINOACYLPROLINE AMINOPEPTIDASE I)		peptidase	1034
1447 8348)	78747988 (8347, Novel Protein sim. GBank g 172061 sp P44817 PEPD_HAEIN - AMINOACYL-HISTIDINE DIPEPTIDASE (XAA- HIS DIPEPTIDASE) (X-HIS DIPEPTIDASE) (BETA-ALANYL-HISTIDINE DIPEPTIDASE) (CARNOSINASE) (PEPTIDASE D)		peptidase	1023
1448 18410)	29342092 (18409, Novel Protein sim. GBank g 127290 sp P10507 MPP2_YEAST - MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR (BETA-MPP) (PEP)		peptidase	1022
1449 20650)	78798792 (20649, Novel Protein sim. GBank g 127290 sp P10507 MPP2_YEAST - MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR (BETA-MPP) (PEP)		peptidase	1022
1450 6992)	82465078 (6991, Novel Protein sim. GBank g 2131729 pi S64780 - hypothetical protein YLL029w - yeast (Saccharomyces cerevisiae)		peptidase	1000, 1001, 1003, 1006, 1008, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1026, 1031, 1034, 1038, 1039, 1044
1451 9796)	8754230 (9795, Novel Protein sim. GBank g 2492531 sp P73971 AMPL_SYNY3 - PUTATIVE CYTOSOL AMINOPEPTIDASE (LEUCINE AMINOPEPTIDASE) (LAP) (LEUCYL AMINOPEPTIDASE)		peptidase	1004

1452	65679586 (12105, 12106)	Novel Protein sim. GBank gi 2492531 sp P73971 AMPL_SYN3 - PUTATIVE CYTOSOL AMINOPEPTIDASE (LEUCINE AMINOPEPTIDASE) (LAP) (LEUCYL AMINOPEPTIDASE E)		peptidase	1016
1453	95002877 (10253, 10254)	Novel Protein sim. GBank gi 2497952 sp P55667 Y4TM_RHISN - HYPOTHETICAL HYDROLASE/PEPTIDASE Y4TM		peptidase	1006
1454	11361003 (9961, 9962)	Novel Protein sim. GBank gi 3859775 emb CA491125.1 - (Z54328) putative aminopeptidase p [Schizosaccharomyces pombe]		peptidase	1024
1455	27839053 (6643, 6644)	Novel Protein sim. GBank gi 461915 sp P33894 DAP1 YEAST - DIPEPTIDYL AMINOPEPTIDASE A (DPAP A) (YSCIV)		peptidase	1006
1456	78462461 (17071, 17072)	Novel Protein sim. GBank gi 4887207 gb AAD32234.1 AF14744 - (AF147448) penicillin-binding protein 5 [Pseudomonas aeruginosa]		peptidase	1008
1457	17696239 (21785, 21786)	Novel Protein sim. GBank gi 9841185 emb CAA62529 - (X91067) 01232 [Saccharomyces cerevisiae]		peptidase	1017
1458	70287106 (18309, 18310)	Novel Protein sim. GBank gi 115722 sp P21179 CATE_ECOLI - CATALASE HPII (HYDROXYPEROXIDASE II)	Contains protein domain (PF00199) - Catalase	peroxidase	1009
1459	78905042 (3145, 3116)	Novel Protein sim. GBank gi 1168785 sp P45739 CATA_HELAN - CATALASE	Contains protein domain (PF00199) - Catalase	peroxidase	1003
1460	78274861 (375, 376)	Novel Protein sim. GBank gi 2493546 sp Q59635 CATB_PSEAE - CATALASE PRECURSOR (PARAQUAT INDUCIBLE CATALASE ISOZYME B)	Contains protein domain (PF00199) - Catalase	peroxidase	1006

1461	79619827 (15319, 15320)	Novel Protein sim. GBank gil2493546 sp Q59635 CATB_PSEAE - CATALASE PRECURSOR (PARAQUAT INDUCIBLE CATALASE ISOZYME B)	Contains protein domain (PF00199) - Catalase	peroxidase	1003
1462	37038208 (15613, 15614)	Novel Protein sim. GBank gil1552875 emb CAB02471 - (2880343) hypothetical protein Rv3792 [Mycobacterium tuberculosis]		peroxidase	1012
1463	10339414 (22279, 22280)	Novel Protein sim. GBank gil687687 (J20224) - similar to catalase, Swiss-Prot Accession Number P04040 [Synechococcus sp.]		peroxidase	1039
1464	80248473 (5787, 5788)	Novel Protein sim. GBank gil130120 sp P23620 PHOB_PSEAE - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATOR PROTEIN PHOB	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	1003, 1004, 1008, 1009, 1010, 1017, 1022, 1034, 1038
1465	79255708 (5761, 5762)	Novel Protein sim. GBank gil1731207 sp Q11156 RGX3_MYCTU - SENSORY TRANSDUCTION PROTEIN REGX3	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	1016
1466	79568139 (10707, 10708)	Novel Protein sim. GBank gil4433393 gb IAAD20723 - (AF051157) response regulator homolog OmpR [Vibrio cholerae]	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	1058
1467	78799611 (677, 678)	Novel Protein sim. GBank gil1709746 sp P53043 PPT1_YEAST - SERINE/THREONINE PROTEIN PHOSPHATASE T (PPT)	Contains protein domain (PF00149) - Ser/Thr protein phosphatase	phosphatase	1009
1468	35605266 (10599, 10600)	Novel Protein sim. GBank gil1723460 sp Q10298 YD44_SCHPO - PUTATIVE SERINE/THREONINE PROTEIN PHOSPHATASE C22H10.04	Contains protein domain (PF00149) - Ser/Thr protein phosphatase	phosphatase	1053
1469	71773894 (17859, 17860)	Novel Protein sim. GBank gil4646217 gb IAD26883.1 AC00729 - (AC00729) putative phosphoprotein phosphatase [Arabidopsis thaliana]	Contains protein domain (PF00149) - Ser/Thr protein phosphatase	phosphatase	1003

1470	39330894 (3665, 3666)	Novel Protein sim. GBank gi 1168743 sp P43165 CAH5_RAT - CARBONIC ANHYDRASE V PRECURSOR (CARBONATE DEHYDRATASE V)	Contains protein domain (PF00194) - Eukaryotic-type carbonic anhydrase	phosphatase	1006, 1022, 1024
1471	37036349 (16525, 16526)	Novel Protein sim. GBank gi 3261659 emb CAB03751 - (Z81368) hypothetical protein Rv2419c [Mycobacterium tuberculosis]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	phosphatase	1012
1472	57147848 (18773, 18774)	Novel Protein sim. GBank gi 3468911 sp P45597 PTF1_XANCP - MULTIPHOSPHORYL TRANSFER PROTEIN (MTP) (CONTAINS: PHOSPHOENOLPYRUVATE:PROTEIN PHOTOTRANSFERASE SYSTEM, ENZYME I); PHOSPHOCARRIER PROTEIN HPR (PROTEIN H); PTS SYSTEM, FRUCTOSE-SPECIFIC II A COMPONENT ...	Contains protein domain (PF00391) - PEP-utilizing enzymes	phosphatase	1004
1473	78378250 (5817, 5818)	Novel Protein sim. GBank gi 2285879 dbj BA21654 - (D86338) pyruvate orthophosphate dikinase [Eleocharis vivipara]	Contains protein domain (PF00391) - PEP-utilizing enzymes	phosphatase	1008
1474	80502627 (9529, 9530)	Novel Protein sim. GBank gi 1781230 emb CAB06277 - (Z83867) hypothetical protein Rv3137 [Mycobacterium tuberculosis]	Contains protein domain (PF00459) - Inositol monophosphatase family	phosphatase	1003, 1009, 1012, 1026, 1034, 1044
1475	80249522 (17093, 17094)	Novel Protein sim. GBank gi 2495224 sp Q06590 HIS7_MYCTU - IMIDAZOGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD)	Contains protein domain (PF00475) - Imidazoleglycerol-phosphate dehydratase	phosphatase	1003, 1006, 1024, 1031
1476	80052549 (18191, 18192)	Novel Protein sim. GBank gi 1154161 sp P23890 CADC_ECOLI - TRANSCRIPTIONAL ACTIVATOR CADC	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1010
1477	37799459 (6997, 6998)	Novel Protein sim. GBank gi 1291581 sp P03025 OMPR_ECOLI - TRANSCRIPTIONAL REGULATORY PROTEIN OMPR	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1012

1478	80027080 (18721, 18722)	Novel Protein sim. GBank gll13012Q1sp1P23620 PHOB_PSEAE - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATOR PROTEIN PHOB	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1006, 1010, 1017, 1024, 1034, 1038
1479	56899823 (6725, 6726)	Novel Protein sim. GBank gll1731207 sp1Q11156 RGX3_MYCTU - SENSORY TRANSDUCTION PROTEIN REGX3	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1003, 1024, 1029, 1053
1480	20294262 (10001, 10002)	Novel Protein sim. GBank gll3236110 emb1 CAA04532 - (AJ001103) arcA [Lactococcus lactis]	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1034
1481	82319797 (3105, 3106)	Novel Protein sim. GBank gll4455703 emb1 CAB36688 - (AL035500) putative two-component response regulator [Mycobacterium leprae]	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1001, 1034
1482	20711865 (17283, 17284)	Novel Protein sim. GBank gll1730805 sp1 P396631 SPHR_SYNTP7 - ALKALINE PHOSPHATASE SYNTHESIS TRANSCRIPTIONAL REGULATOR PROTEIN SPHR	Contains protein domain (PF00486) - Transcriptional regulatory protein, C terminal	phosphatase	1022
1483	29514479 (197, 198)	Novel Protein sim. GBank gll396351 U00006 - isocitrate dehydrogenase kinase/phosphatase [Escherichia coli]	Contains protein domain (PF00633) - Helix-hairpin- helix motif.	phosphatase	1024
1484	78793915 (2977, 2978)	Novel Protein sim. GBank gll10844555 pir S56280 - histidinol-phosphatase (EC 3.1.3.15) - yeast (Saccharomyces cerevisiae)		phosphatase	1028
1485	11077119 (15511, 15512)	Novel Protein sim. GBank gll1172570 sp1 Q05205 PPB_LYSEN - ALKALINE PHOSPHATASE PRECURSOR (APASE)		phosphatase	1022

1486	80453168 (327, 328)	Novel Protein sim. GBank gi 1173429 sp P42941 SERB_YEAST - PHOSPHOSERINE PHOSPHATASE (PSP) (O-PHOSPHOSERINE PHOSPHOHYDROLASE) (PSP)		phosphatase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1022, 1024, 1027, 1031, 1034, 1038, 1039, 1044
1487	80217473 (4807, 4808)	Novel Protein sim. GBank gi 131557 sp P25044 PTP1_YEAST - PROTEIN-TYROSINE PHOSPHATASE 1 (PTPASE 1)		phosphatase	1017, 1022, 1024
1488	80216827 (11995, 11996)	Novel Protein sim. GBank gi 131557 sp P25044 PTP1_YEAST - PROTEIN-TYROSINE PHOSPHATASE 1 (PTPASE 1)		phosphatase	1004, 1006, 1010, 1022, 1023, 1024, 1026, 1031, 1034
1489	27261124 (7143, 7144)	Novel Protein sim. GBank gi 1622933 U72346 protein phosphatase type 2C [Saccharomyces cerevisiae]		phosphatase	1022
1490	78470741 (3131, 3132)	Novel Protein sim. GBank gi 2160692 U73527 B' regulatory subunit of PP2A [Arabidopsis thaliana]		phosphatase	1053
1491	84453007 (20393, 20394)	Novel Protein sim. GBank gi 2499743 sp Q06216 PIG1_YEAST - PROTEIN PHOSPHATASE 1 REGULATORY SUBUNIT PIG1		phosphatase	1001, 1003, 1006, 1009, 1010, 1012, 1022, 1023, 1024, 1026, 1027, 1031, 1034, 1039, 1044, 1049
1492	29250109 (8477, 8478)	Novel Protein sim. GBank gi 2550891 bb s108080 - (S39392) protein tyrosine phosphatase, PTase [EC 3.1.3.48] [human, colon, Peptide Partial, 704 aa] [Homo sapiens]		phosphatase	1026
1493	79600986 (10497, 10498)	Novel Protein sim. GBank gi 2626321 dbj BAA23400 - (D83967) YfkJ [Bacillus subtilis]		phosphatase	1021, 1022, 1039

1494	20446470 (15639, 15640)	Novel Protein sim. G-Bank gi 266860 sp P2946 PTP2_YEAST - PROTEIN-TYROSINE PHOSPHATASE 2 (PTPASE 2)		phosphatase	1024
1495	70957975 (5855, 5856)	Novel Protein sim. G-Bank gi 2791640 emb CAA16127.1 - (AL021287) serB2 [Mycobacterium tuberculosis]		phosphatase	1029
1496	29906853 (18867, 18868)	Novel Protein sim. G-Bank gi 2828820 (AF012836) - trehalose/maltose binding protein [Thermococcus litoralis]		phosphatase	1024
1497	39564625 (20227, 20228)	Novel Protein sim. G-Bank gi 2851536 sp P11491 PPB_YEAST - REPRESSIBLE ALKALINE PHOSPHATASE PRECURSOR		phosphatase	1010
1498	25258660 (16395, 16396)	Novel Protein sim. G-Bank gi 2944180 (AF007779) - trehalose-6-phosphate phosphatase [Arabidopsis thaliana]		phosphatase	1008
1499	80034053 (17571, 17572)	Novel Protein sim. G-Bank gi 3127849 emb CAA18915 - (AL023496) hypothetical protein [Streptomyces coelicolor]		phosphatase	1006, 1016, 1029
1500	87456552 (15039, 15040)	Novel Protein sim. G-Bank gi 5525080 emb CA550876.1 - (AL096844) putative phosphoserine phosphatase [Streptomyces coelicolor A3(2)]		phosphatase	1015, 1029
1501	78368938 (151, 152)	Novel Protein sim. G-Bank gi 586390 sp P38148 YB9T_YEAST - PROBABLE PROTEIN-TYROSINE PHOSPHATASE YBR276C		phosphatase	1009
1502	33198871 (19517, 19518)	Novel Protein sim. G-Bank gi 586390 sp P38148 YB9T_YEAST - PROBABLE PROTEIN-TYROSINE PHOSPHATASE YBR276C		phosphatase	1029
1503	20731437 (16087, 16088)	Novel Protein sim. G-Bank gi 729359 sp P38774 DOG1_YEAST - 2-DEOXYGLUCOSE-6-PHOSPHATE PHOSPHATASE 1 (2-DOG-6-P 1) (2-DEOXYGLUCOSE-6-PHOSPHATASE 1)		phosphatase	1003, 1006

1504	80250205 (17437, 17438)	Novel Protein sim. GBank g 731885 sp P40454 YIP3 - YEAST - HYPOTHETICAL 45.1 KD PROTEIN IN IMP2-DNA43 INTERGENIC REGION		phosphatase	1006
1505	11034025 (3487, 3488)	Novel Protein sim. GBank g 90254 sp I(A28334 - protein-tyrosine-phosphatase (EC 3.1.3.48) Ly-5 precursor. (B-cell variant) - mouse		phosphatase	1001
1506	17890745 (22041, 22042)	Novel Protein sim. GBank g 341341 emb CAA20272 - (AL031231) guanosine pentaphosphate synthetase/polyribonucleotide nucleotidyltransferase [Streptomyces coelicolor]	Contains protein domain (PF00013) - KH domain	phosphorylase	1039
1507	95294456 (7257, 7258)	Novel Protein sim. GBank g 341341 emb CAA20272 - (AL031231) guanosine pentaphosphate synthetase/polyribonucleotide nucleotidyltransferase [Streptomyces coelicolor]	Contains protein domain (PF00013) - KH domain	phosphorylase	1001, 1006, 1009, 1010, 1022, 1023, 1025, 1031, 1034, 1038
1508	66396759 (22247, 22248)	Novel Protein sim. GBank g 2645846 (AF033856) - glycogen phosphorylase [Agrobacterium tumefaciens]	Contains protein domain (PF00343) - Carbohydrate Phosphorylases	phosphorylase	1008
1509	94684418 (11129, 11130)	Novel Protein sim. GBank g 606352 (U18997) - maltodextrin phosphorylase [Escherichia coli]	Contains protein domain (PF00343) - Carbohydrate Phosphorylases	phosphorylase	1024
1510	561162488 (18121, 18122)	Novel Protein sim. GBank g 7288 emb CAA44069 - (X62142) glycogen phosphorylase 1 [Dictyostelium discoideum]	Contains protein domain (PF00343) - Carbohydrate Phosphorylases	phosphorylase	1003, 1017
1511	56025787 (7651, 7652)	Novel Protein sim. GBank g 1172546 sp P44584 PNP_HAEIN - POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE (POLYNUCLEOTIDE PHOSPHORYLASE) (PNPASE)		phosphorylase	1003
1512	11092964 (20311, 20312)	Novel Protein sim. GBank g 2462716 emb CAA71697 - (Y10692) ppn [Yersinia enterocolitica]		phosphorylase	1006
1513	11265627 (20689, 20690)	Novel Protein sim. GBank g 2506345 sp P45563 XAPA_ECOLI - XANTHOSINE PHOSPHORYLASE		phosphorylase	1022

1514	78083034 (4061, 4062)	Novel Protein sim. GBank gi 3413411 emb CAA20272 - (AL031231) guanosine pentaphosphate synthetase/ polyribonucleotide nucleotidyltransferase/ [Streptomyces coelicolor]		phosphorylase	1008, 1029
1515	46872595 (11413, 11414)	Novel Protein sim. GBank gi 4389235 pdbl BRW A - Chain A, The Crystal Structure Of Pyrimidine Nucleoside Phosphorylase In A Closed Conformation		phosphorylase	1029
1516	37025276 (19011, 19012)	Novel Protein sim. GBank gi 642066 (U19858) - guanosine pentaphosphate synthetase [Streptomyces antibioticus]		phosphorylase	1038
1517	78737112 (5335, 5336)	Novel Protein sim. GBank gi 1724341 sp P39684 PES4_YEAST - PES4 PROTEIN (DNA POLYMERASE EPSILON SUPPRESSOR 4)	Contains protein domain (PF00018) - SH3 domain	polymerase	1003, 1022, 1023, 1026, 1039, 1044
1518	79797949 (13121, 13122)	Novel Protein sim. GBank gi 1737492 (U81318) - poly(A)-binding protein [Triticum aestivum]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	polymerase	1009, 1044
1519	78785762 (12067, 12068)	Novel Protein sim. GBank gi 417441 sp P04147 PABP_YEAST - POLYADENYLATE BINDING PROTEIN, CYTOPLASMIC AND NUCLEAR (PABP) (ARS CONSENSUS BINDING PROTEIN ACBP-57) (POLYADENYLATE TAIL-BINDING PROTEIN)	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	polymerase	1008, 1017
1520	80439935 (15283, 15284)	Novel Protein sim. GBank gi 1189011 sp P14284 DPOZ_YEAST - DNA POLYMERASE ZETA CATALYTIC SUBUNIT	Contains protein domain (PF00136) - DNA polymerase family B	polymerase	1004, 1008, 1012, 1022, 1024, 1031, 1038
1521	79255837 (12811, 12812)	Novel Protein sim. GBank gi 172209 (J03268) - polymerase I [Saccharomyces cerevisiae]	Contains protein domain (PF00136) - DNA polymerase family B	polymerase	1003

1522	85513835 (1367, 1368)	Novel Protein sim. GBank gi 1168816 sp P30665 CC54_YEAST - CELL DIVISION CONTROL PROTEIN 54	Contains protein domain (PF00493) - MCM2/3/5 family	polymerase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1016, 1017, 1022, 1023, 1024, 1025, 1026, 1034, 1039, 1044, 1049
1523	65470792 (8385, 8386)	Novel Protein sim. GBank gi 4808392 emb CAB42723.1 - (ALD049826) putative RNA nucleotidyltransferase [Streptomyces coelicolor]	Contains protein domain (PF01966) - HD domain	polymerase	1049
1524	78935955 (3469, 3470)	Novel Protein sim. GBank gi 1168816 sp P30665 CC54_YEAST - CELL DIVISION CONTROL PROTEIN 54		polymerase	1003, 1004, 1044
1525	78385332 (683, 684)	Novel Protein sim. GBank gi 1175777 sp P441176 YCIV_HAEIN - HYPOTHETICAL PROTEIN H11400		polymerase	1003, 1016, 1026, 1030, 1053
1526	13076416 (1997, 1998)	Novel Protein sim. GBank gi 118794 sp P10443 DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	1027
1527	80062943 (881, 882)	Novel Protein sim. GBank gi 118804 sp P27903 DP3B_STRCO - DNA POLYMERASE III, BETA CHAIN		polymerase	1010, 1024
1528	27966085 (15595, 15596)	Novel Protein sim. GBank gi 118841 sp P21951 DPOE_YEAST - DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (DNA POLYMERASE II SUBUNIT A)		polymerase	1034
1529	80503756 (21495, 21496)	Novel Protein sim. GBank gi 118841 sp P21951 DPOE_YEAST - DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (DNA POLYMERASE II SUBUNIT A)		polymerase	1003, 1006, 1010, 1012, 1017, 1022, 1023, 1026, 1027, 1034, 1038, 1039, 1044

1530	80242388 (18945, 18946)	Novel Protein sim. GBank gi 118901 sp P14284 DPOZ_YEAST - DNA POLYMERASE ZETA CATALYTIC SUBUNIT		polymerase	1003, 1008, 1022, 1025
1531	82059356 (12585, 12586)	Novel Protein sim. GBank gi 172209 (J03268) - polymerase I [Saccharomyces cerevisiae]		polymerase	1003, 1004, 1006, 1008, 1010, 1016, 1022, 1024, 1026, 1034
1532	79113145 (14067, 14068)	Novel Protein sim. GBank gi 172209 (J03268) - polymerase I [Saccharomyces cerevisiae]		polymerase	1039
1533	39564426 (15983, 15984)	Novel Protein sim. GBank gi 172209 (J03268) - polymerase I [Saccharomyces cerevisiae]		polymerase	1004, 1006, 1010, 1022, 1034
1534	80237518 (18757, 18758)	Novel Protein sim. GBank gi 2105050 emb CAB08836 - (Z95436) hypothetical protein Rv3644c [Mycobacterium tuberculosis]		polymerase	1006, 1023, 1031
1535	12929511 (11675, 11676)	Novel Protein sim. GBank gi 2293260 (AF008220) - DNA-polymerase III alpha-chain [Bacillus subtilis]		polymerase	1024
1536	80219092 (2461, 2462)	Novel Protein sim. GBank gi 2311717 sp P29496 MCM5_YEAST - MINICHROMOSOME MAINTENANCE PROTEIN 5 (CELL DIVISION CONTROL PROTEIN 46)		polymerase	1004, 1006, 1010, 1022, 1024, 1034, 1044
1537	80081373 (12457, 12458)	Novel Protein sim. GBank gi 2311717 sp P29496 MCM5_YEAST - MINICHROMOSOME MAINTENANCE PROTEIN 5 (CELL DIVISION CONTROL PROTEIN 46)		polymerase	1004, 1022
1538	377814367 (20523, 20524)	Novel Protein sim. GBank gi 2497820 sp Q43704 MCM3_MAIZE - DNA REPLICATION LICENSING FACTOR MCM3 HOMOLOG (REPLICATION ORIGIN ACTIVATOR) (ROA PROTEIN)		polymerase	1058
1539	91006450 (3537, 3538)	Novel Protein sim. GBank gi 2661649 emb CAA15755 - (AL009198) dnaE2 [Mycobacterium tuberculosis]		polymerase	1020

1540	16485432 (3747, 3748)	Novel Protein sim. GBank gi 2661649 emb CAA15755 - (AL009198) dnaE2 [Mycobacterium tuberculosis]		polymerase	1008
1541	79471264 (9361, 9362)	Novel Protein sim. GBank gi 2661649 emb CAA15755 - (AL009198) dnaE2 [Mycobacterium tuberculosis]		polymerase	1022
1542	11103311 (10491, 10492)	Novel Protein sim. GBank gi 2960145 emb CAA18043.1 - (AL022121) dnazX [Mycobacterium tuberculosis]		polymerase	1004
1543	20620086 (11403, 11404)	Novel Protein sim. GBank gi 3150103 emb CAA19155 - (AL023596) DNA polymerase III subunits gamma and tau [Mycobacterium leprae]		polymerase	1004
1544	20428686 (21967, 21968)	Novel Protein sim. GBank gi 3860586 emb CAA14487 - (AJ235270) POLY(A) POLYMERASE (pcnB) [Rickettsia prowazekii]		polymerase	1022
1545	10250675 (14061, 14062)	Novel Protein sim. GBank gi 3915668 sp P26464 DP3B_SALTY - DNA POLYMERASE III, BETA CHAIN		polymerase	1038
1546	79588540 (22913, 22914)	Novel Protein sim. GBank gi 4324655 gb AAD16978 - (AF108191) DNA polymerase III alpha subunit [Streptomyces coelicolor]		polymerase	1017, 1030, 1034
1547	37808879 (22801, 22802)	Novel Protein sim. GBank gi 5590665 L37060 - putative Murine leukemia virus]		polymerase	1058
1548	8498578 (5705, 5706)	Novel Protein sim. GBank gi 585063 sp P38121 DPO2_YEAST - DNA POLYMERASE ALPHA/PRIMASE ASSOCIATED SUBUNIT (P86 SUBUNIT)		polymerase	1004
1549	789955058 (13027, 13028)	Novel Protein sim. GBank gi 5866518 sp P38251 RFC5_YEAST - ACTIVATOR 1 SUBUNIT 5 (REPLICATION FACTOR C SUBUNIT 5)		polymerase	1003

1550	99444632 (17121, 17122)	Novel Protein sim. GBank gi 2832781 emb CAAA12645 - (A 225805 inward potassium channel alpha subunit [Eugenia densa]	Contains protein domain (PF00023) - Ank repeat	potassium_channel	1001, 1004, 1009, 1010, 1022, 1037, 1049, 1054
1551	30804885 (16809, 16810)	Novel Protein sim. GBank gi 3641300 (AF087453) - potassium channel [Rattus norvegicus]		potassium_channel	1026
1552	16316084 (6337, 6338)	Novel Protein sim. GBank gi 4826790 ref NP_004969.1 pkCNC - potassium voltage-gated channel, Shaw-related subfamily, member 4.		potassium_channel	1054
1553	86685963 (14323, 14324)	Novel Protein sim. GBank gi 4504627 ref NP_001533.1 plGSF - immunoglobulin superfamily, member 3		prostaglandin	1022, 1024, 1026
1554	88097674 (4945, 4946)	Novel Protein sim. GBank gi 2495021 sp Q63645 PAR2_RAT - PROTEINASE ACTIVATED RECEPTOR 2 PRECURSOR (PAR-2)	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	protease	1022, 1025
1555	87453833 (12053, 12054)	Novel Protein sim. GBank gi 1168536 sp P42210 ASPR_HORVU - PHYTEPSIN PRECURSOR (ASPARTIC PROTEINASE)	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	protease	1007, 1025
1556	80493897 (2369, 2370)	Novel Protein sim. GBank gi 1709041 sp P53379 MKC7_YEAST - ASPARTIC PROTEINASE MKC7 PRECURSOR	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	protease	1001, 1004, 1012, 1017, 1022, 1025, 1034
1557	80236862 (5771, 5772)	Novel Protein sim. GBank gi 250724 sp P32339 YAP3_YEAST - ASPARTIC PROTEINASE 3 PRECURSOR (YAPSIN 1)	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	protease	1003, 1006, 1008, 1017, 1022, 1024, 1026, 1038
1558	27823866 (19967, 19968)	Novel Protein sim. GBank gi 500621 (L31651) - aspartyl protease 3 [Saccharomyces cerevisiae]	Contains protein domain (PF00026) - Eukaryotic aspartyl protease	protease	1001, 1006, 1010, 1017, 1022, 1024, 1044
1559	80084513 (16273, 16274)	Novel Protein sim. GBank gi 4218542 emb CAAO9631.1 - (AJ011500) putative two-component response regulator [Streptomyces violaceoruber]	Contains protein domain (PF00072) - Response regulator receiver domain	protease	1034

1560	79834209 (15219, 15220)	Novel Protein sim. GBank g 505140 emb CAB4989.1 - (AJ131854) NarL protein [Pseudomonas stutzeri]	Contains protein domain (PF00072) - Response regulator receiver domain	protease	1008
1561	80430995 (5153, 5154)	Novel Protein sim. GBank g 140497 sp P253B1 YCT5_YEAST - HYPOTHETICAL SUBTILASE-TYPE PROTEINASE PRECURSOR IN TSM1-ARE1 INTERGENIC REGION	Contains protein domain (PF00082) - Subtilase family	protease	1004, 1006, 1008, 1009, 1010, 1012, 1017, 1022, 1024, 1026, 1031, 1038, 1039, 1054
1562	8504426 (7591, 7592)	Novel Protein sim. GBank g 4506153 ref NP_002784.1 pPRSS - protease, serine, 8 (prostasin)	Contains protein domain (PF00089) - Trypsin	protease	1004
1563	29512480 (10071, 10072)	Novel Protein sim. GBank g 543374 pir JC2125 - chymase (EC 3.4.21.39) precursor - rat	Contains protein domain (PF00089) - Trypsin	protease	1016
1564	78758112 (13703, 13704)	Novel Protein sim. GBank g 1046373 (U37336) - SAG12 protein [Arabidopsis thaliana]	Contains protein domain (PF00112) - Papain family cysteine protease	protease	1039
1565	20467281 (10799, 10800)	Novel Protein sim. GBank g 3581871 emb CAA20827 - (AL031541) putative transcriptional regulator [Streptomyces coelicolor]	Contains protein domain (PF00126) - Bacterial regulatory helix-turn-helix protein, lysR family	protease	1010
1566	10126494 (12075, 12076)	Novel Protein sim. GBank g 4063015 (AF083061) - protease PtaA [Pseudomonas fluorescens]	Contains protein domain (PF00353) - Hemolysin-type calcium-binding proteins	protease	1017
1567	71730420 (15271, 15272)	Novel Protein sim. GBank g 3834446 gb AAC70948.1 - (AF071885) ATP-dependent Clp protease proteolytic subunit 2 [Streptomyces coelicolor]	Contains protein domain (PF00574) - Clp protease	protease	1008
1568	65676443 (14793, 14794)	Novel Protein sim. GBank g 1345104 (U30799) - MucD [Azotobacter vinelandii]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	protease	1046

1569	38868615 (21627, 21628)	Novel Protein sim. GBank g 3193167 (AF067834) - caspase-8 [Mus musculus]	Contains protein domain (PF00655) - ICE-like protease (caspase) p10	protease	1022, 1039
1570	94747729 (7157, 7158)	Novel Protein sim. GBank g 4506221 refNP_002807.1 pPSMD - proteasome (prosome, macropain) 26S subunit, non-ATPase, 12.	Contains protein domain (PF01399) - PCI domain	protease	1003
1571	30236456 (15573, 15574)	Novel Protein sim. GBank g j283937 pir S28258 - androgen-regulated epididymal protein precursor - crab-eating macaque	Contains protein domain (PF01562) - Reprolysin family propeptidase	protease	1030
1572	26255842 (10863, 10864)	Novel Protein sim. GBank g j1147800 (U43720) - Sug2p [Saccharomyces cerevisiae]		protease	1003, 1017, 1022, 1024, 1034
1573	94712754 (10633, 10634)	Novel Protein sim. GBank g j1174631 sp P46465 PRSAORYSA - 26S PROTEASE REGULATORY SUBUNIT 6A HOMOLOG (TAT-BINDING PROTEIN HOMOLOG 1),[TBP-1],		protease	1010
1574	78941973 (489, 470)	Novel Protein sim. GBank g j1175759 sp P45181 PQQQ_HAEIN - PROBABLE ZINC PROTEASE PQQL		protease	1023
1575	11286841 (21263, 21264)	Novel Protein sim. GBank g j125350 sp P13134 KEX2_YEAST - KEXIN PRECURSOR (KEX2 PROTEASE) (PROTEINASE YSCF)		protease	1006
1576	30170092 (65, 66)	Novel Protein sim. GBank g j1296452 emb CAA64618 - (X95306) CipX protein [Bacillus subtilis]		protease	1030
1577	38890199 (18953, 18954)	Novel Protein sim. GBank g j1352918 sp P47154 ST24_YEAST - CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFIC ENDOPROTEASE 1) (PPSEEP 1) (A-FACTOR CONVERTING ENZYME)		protease	1008

1578	37800133 (5499, 5500)	Novel Protein sim. GBank gil1653007[db]BAA17924 - (D90910) rubisco operon transcriptional regulator [Synechocystis sp.]		protease	1012
1579	80235198 (12841, 12842)	Novel Protein sim. GBank gil1930112 (U94710) - type I procollagen C-proteinase enhancer protein [Rattus norvegicus]		protease	1004, 1008, 1010, 1022, 1024, 1044
1580	87914009 (16379, 16380)	Novel Protein sim. GBank gil2120522[pir][PC6018] ATP-dependent Clp proteinase homolog - Azospirillum brasiliense (fragment)		protease	1015
1581	17106023 (3783, 3784)	Novel Protein sim. GBank gil2131346[pir][S87695] hypothetical protein YDL147w - yeast (Saccharomyces cerevisiae)		protease	1008
1582	79606717 (20855, 20856)	Novel Protein sim. GBank gil2132507[pir][S861146] probable membrane protein YDR349c - yeast (Saccharomyces cerevisiae)		protease	1006, 1029, 1037
1583	78746498 (13969, 13970)	Novel Protein sim. GBank gil2492519[sp]Q41365 PRS7 SPIOL - 26S PROTEASE REGULATORY SUBUNIT 7 (26S PROTEASOME SUBUNIT 7)		protease	1008
1584	29687208 (3911, 3912)	Novel Protein sim. GBank gil4049891[sp]P76403 YEGQ_ECOLI - PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION		protease	1016
1585	788906950 (4859, 4860)	Novel Protein sim. GBank gil2499891[sp]P76403 YEGQ_ECOLI - PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION		protease	1023
1586	27977763 (13825, 13826)	Novel Protein sim. GBank gil2791497[emb]CAA16034 - (A021246) clpX [Mycobacterium tuberculosis]		protease	1006
1587	949911719 (2563, 2564)	Novel Protein sim. GBank gil2879941[db]BA-A24832 - (AB010960) MIFR [Rattus norvegicus]		protease	1004

1588	87454361 (175, 176)	Novel Protein sim. GBank gil29602[6]embICAA11285] - (AJ223384) 26S proteasome regulatory ATPase subunit 10b (S10b) [Manduca sexta]		protease	1025
1589	87454367 (833, 834)	Novel Protein sim. GBank gil29602[6]embICAA11285] - (AJ223384) 26S proteasome regulatory ATPase subunit 10b (S10b) [Manduca sexta]		protease	1025
1590	95009187 (1653, 1654)	Novel Protein sim. GBank gil2984005 (AE000750) - ATPase subunit of ATP-dependent protease [Aequifex aequipinnatus]		protease	1003
1591	33206121 (19521, 19522)	Novel Protein sim. GBank gil304908 (L20572) - ATP-dependent protease [Escherichia coli]		protease	1026
1592	6748992 (14443, 14444)	Novel Protein sim. GBank gil3115348 (AC004668) - mitochondrial processing peptidase beta subunit precursor; similar to Q03346 (PID:g1171010) [Homo sapiens]		protease	1022
1593	20729986 (21895, 21896)	Novel Protein sim. GBank gil3913521[sp O66592 DUT_AQUAE - DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (DUTPase) (DUTP PYROPHOSPHATASE)		protease	1006
1594	78895306 (4103, 4104)	Novel Protein sim. GBank gil3913944[sp P74936 LON_VIBPA - ATP-DEPENDENT PROTEASE LA		protease	1003, 1017, 1022, 1039
1595	79485827 (3551, 3552)	Novel Protein sim. GBank gil3913995[sp P77810 LON_AZOBR - ATP-DEPENDENT PROTEASE LA		protease	1022
1596	13883828 (13887, 13888)	Novel Protein sim. GBank gil4154933 (AE001474) - PROTEASE DO [Helicobacter pylori J99]		protease	1053
1597	80089314 (18933, 18934)	Novel Protein sim. GBank gil4416302[gb AAD20307] - (AF105716) copia-type polyprotein [Zea mays]		protease	1010
1598	34715608 (19115, 19116)	Novel Protein sim. GBank gil4416302[gb AAD20307] - (AF105716) copia-type polyprotein [Zea mays]		protease	1016

1599 5502)	17952065 (5501, Novel Protein sim. GBank g I450614 ref NP_002766.1 pPRSS - protease, serine, 1 (IGF binding)		protease	1023
1600 5474)	Novel Protein sim. GBank g I4506213 ref NP_002796.1 pPSMC - proteasome (prosome, macropain) 26S subunit, ATPase, 5		protease	1006
1601 3240)	Novel Protein sim. GBank g I454438 (L28110) - LON gene of <i>S. cerevisiae</i> is downstream of the HAP 3 gene; Putative ATP-binding motif bp 1960 to bp 1986.; Putative catalytic site serine of serine proteases from bp 3109 to bp 3111 [<i>Saccharomyces cerevisiae</i>]		protease	1010, 1024
1602 16304)	Novel Protein sim. GBank g I4587771 gb AAD25872.1 AF02750 - (AF027500) ATP-dependent Clp protease regulatory subunit [Aquitex pyrophilus]		protease	1008
1603 17838)	Novel Protein sim. GBank g I461753 sp P35100 CLPA_PEA - ATP- DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA HOMOLOG PRECURSOR		protease	1044
1604 10644)	Novel Protein sim. GBank g I464486 sp P33298 PRS6_YEAST - 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG (YNT1 PROTEIN) (TAT-BINDING HOMOLOG 2)		protease	1003, 1009, 1017, 1039
1605 17316)	Novel Protein sim. GBank g I549612 sp P36135 YK22_YEAST - HYPOTHETICAL 46.9 KD PROTEIN IN GAP1- NAP1 INTERGENIC REGION		protease	1001, 1004, 1006, 1010, 1017, 1022, 1023, 1024, 1026, 1034, 1044
1606 18526)	Novel Protein sim. GBank g I557599 (L17040) - ATPase [<i>Saccharomyces cerevisiae</i>]		protease	1026
1607 9410)	Novel Proteih sim. GBank g I586344 sp P38323 YB77_YEAST - HYPOTHETICAL 57.9 KD PROTEIN IN PDB1- ABD1 INTERGENIC REGION		protease	1003

1608	78747819 (7355, 7356)	Novel Protein sim. GBank gil 72893 isp P40851 AXL1_YEAST - PUTATIVE PROTEASE AXL1		protease	1022
1609	57293687 (7405, 7406)	Novel Protein sim. GBank gil 72893 isp P40851 AXL1_YEAST - PUTATIVE PROTEASE AXL1		protease	1006, 1008, 1023, 1034
1610	71775532 (2169, 2170)			protease	1003
1611	56146622 (8183, 8184)	Novel Protein sim. GBank gil 772998 emb CAAA71412 - (Y10386) C1 inhibitor [Mus musculus]	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	proteaseinhib	1022
1612	79377566 (5375, 5376)	Novel Protein sim. GBank gil 203063 (M32247) - alpha-1-antitrypsin precursor [Rattus norvegicus]	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	proteaseinhib	1022, 1029
1613	57290025 (14341, 14342)	Novel Protein sim. GBank gil 220649 dbj BAA00579 - (D00675) precursor of alpha-1-protease inhibitor [Rattus norvegicus]	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	proteaseinhib	1004, 1006, 1010, 1022, 1024, 1034
1614	29359122 (21021, 21022)	Novel Protein sim. GBank gil 2507337 sp P05544 CPI3_RAT - CONTRAPSIN LIKE PROTEASE INHIBITOR_3 PRECURSOR (CPI-23) (SERINE PROTEASE INHIBITOR 1) (SPI-1)	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	proteaseinhib	1006, 1022, 1024
1615	17873916 (18695, 18896)	Novel Protein sim. GBank gil 2842653 sp Q62477 WDNM1 PROTEIN PRECURSOR	Contains protein domain (PF00095) - WAP-type (Whey Acidic Protein) 'four-disulfide core'	proteaseinhib	1003
1616	86474895 (10403, 10404)	Novel Protein sim. GBank gil 4038038 (AC005936) - putative proteinase inhibitor II [Arabidopsis thaliana]	Contains protein domain (PF00304) - Gamma-thionins family	proteaseinhib	1038
1617	78720669 (15166)	Novel Protein sim. GBank gil 12893 sp P14046 A13_RAT - ALPHA-1-INHIBITOR III PRECURSOR	Contains protein domain (PF01835) - Alpha-2-macroglobulin family N-terminal region	proteaseinhib	1022, 1026

1618 9966)	28387472 (9967, Novel Protein sim. GBank g 91949 pir S75904 - alpha-1 proteinase inhibitor III, variant 1 precursor rat	Contains protein domain (PF01835) - Alpha-2- macroglobulin family N- terminal region	proteaseinhib	1017
1619 11532)	17684363 (11531, Novel Protein sim. GBank g 1772998 emb CAA71412 - (Y10386) C1 Inhibitor [Mus musculus]		proteaseinhib	1017
1620 5100)	25145622 (5099, Novel Protein sim. GBank g 205089 (M11885) - LMW T-kininogen II precursor [Rattus norvegicus]		proteaseinhib	1003
1621 7128)	27992015 (7127, Novel Protein sim. GBank g 2995378 emb CAA49540 - (X69915) unnamed protein product [Triticum aestivum]	Contains protein domain (PF00085) - Thioredoxin	reductase	1017
1622 12158)	91232617 (12157, Novel Protein sim. GBank g 3122936 sp O30974 THIO_MYCSM - THIOREDOXIN (TRX)	Contains protein domain (PF00085) - Thioredoxin	reductase	1001, 1003, 1006, 1017, 1024, 1026, 1044
1623 910)	70957947 (909, Novel Protein sim. GBank g 5525057 emb CAB50883.1 - (AL096844) probable 3-oxacyl-(acyl-carrier-protein) reductase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF00106) - short chain dehydrogenase	reductase	1029
1624 4606)	11130562 (4605, Novel Protein sim. GBank g 118974 sp P00379 DYR_ECOLI - DIHYDROFOLATE REDUCTASE	Contains protein domain (PF00186) - Dihydrofolate reductase	reductase	1006
1625 8492)	78447989 (8491, Novel Protein sim. GBank g 102057 pir A32950 - probable reductase protein - Leishmania major	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1003
1626 19250)	20384761 (19249, Novel Protein sim. GBank g 102057 pir A32950 - probable reductase protein - Leishmania major	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1004
1627 8472)	80429754 (8471, Novel Protein sim. GBank g 1176540 sp P47137 YJ66 YEAST - PROBABLE OXIDOREDUCTASE YJR098V	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1001, 1006, 1012, 1017, 1022, 1023, 1024, 1027, 1031, 1034, 1038, 1039, 1044

1628	82313706 (20675, 20676)	Novel Protein sim. GBank gi 1176340 sp P47137 YJ866_YEAST - PROBABLE OXIDOREDUCTASE YJR096W	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1001, 1003, 1004, 1006, 1008, 1010, 1012, 1022, 1023, 1024, 1027, 1031, 1034, 1038, 1054
1629	82619195 (19859, 19860)	Novel Protein sim. GBank gi 2492802 sp Q12458 YPR1_YEAST - PUTATIVE REDUCTASE 1	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1006, 1010, 1022, 1024, 1034
1630	820662057 (7057, 7058)	Novel Protein sim. GBank gi 4007669 emb CA22355 - (AL034443) putative oxidoreductase [Streptomyces coelicolor]	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1016, 1017, 1022, 1034
1631	52559923 (22559, 22560)	Novel Protein sim. GBank gi 4895205 gb AAD32792.1 AC00766 - (AC007661) putative alcohol dehydrogenase [Arabidopsis thaliana]	Contains protein domain (PF00248) - Aldo/keto reductase family	reductase	1022
1632	33708560 (841, 842)	Novel Protein sim. GBank gi 2500207 sp P55932 RIR1_HELPY - RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN (RIBONUCLEOTIDE REDUCTASE)	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1016
1633	11085993 (12977, 12978)	Novel Protein sim. GBank gi 3077613 emb CAA70765 - (Y09572) ribonucleotide reductase subunit R1E [Corynebacterium ammoniagenes]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1010
1634	80496353 (16247, 16248)	Novel Protein sim. GBank gi 3402271 emb CAA12167 - (AJ224870) vitamin B12-dependent ribonucleotide reductase [Streptomyces clavuligerus]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1003, 1006, 1010, 1012, 1024, 1026, 1034
1635	13085468 (2941, 2942)	Novel Protein sim. GBank gi 4377312 gb AAD19121 - (AE001679) Ribonucleoside Reductase, Large Chain [Chlamydia pneumoniae]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1027

1636	20724399 (3721, 3722)	Novel Protein sim. GBank gi 4377312 gb [AAD19121] - (AE001679) Ribonucleoside Reductase, Large Chain [Chlamydia pneumoniae]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1006
1637	10245731 (14905, 14906)	Novel Protein sim. GBank gi 4490609 emb [CAB38642.1] - (AJ133495) ribonucleotide reductase major subunit [Staphylococcus aureus]	Contains protein domain (PF00317) - Ribonucleotide reductase	reductase	1022
1638	11077823 (6627, 6628)	Novel Protein sim. GBank gi 3929336 sp Q31214 UCRL_CHRV1 - UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT (RIESKE IRON- SULFUR PROTEIN) (RISP)	Contains protein domain (PF00355) - Rieske [2Fe-2S] domain	reductase	1024
1639	20725553 (10717, 10718)	Novel Protein sim. GBank gi 3560010 emb [CAA20632] - (AL031615) putative nitrate reductase alpha chain [Streptomyces coelicolor]	Contains protein domain (PF00384) - Prokaryotic molybdopterin oxidoreductases	reductase	1006
1640	78235957 (20981, 20982)	Novel Protein sim. GBank gi 1706756 sp P55336 FABG_VIBHA - 3- OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE)	Contains protein domain (PF00678) - Short chain dehydrogenase/reductase C- terminus	reductase	1050
1641	15024237 (19045, 19046)	Novel Protein sim. GBank gi 120483 sp P00363 FRDA_ECOLI - FUMARATE REDUCTASE FLAVOPROTEIN SUBUNIT	Contains protein domain (PF00890) - FAD binding domain	reductase	1054
1642	16456622 (20427, 20428)	Novel Protein sim. GBank gi 1073085 pir S53413 - probable membrane protein YLR426w - yeast (Saccharomyces cerevisiae)		reductase	1008
1643	80079318 (951, 952)	Novel Protein sim. GBank gi 1169220 sp P46839 DAPB_MYCBO - DIHYDRODIPICOLINATE REDUCTASE		reductase	1034

1644	78749596 (11705, 11706)	Novel Protein sim. GBank gi 1169360 sp P45004 DMSA_HAEIN - ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN A PRECURSOR (DMSO REDUCTASE)		reductase	1022
1645	11810208 (8825, 8826)	Novel Protein sim. GBank gi 1176118 sp P42593 FADH_ECOLI - 2,4- DIENOYL-COA REDUCTASE (NADPH) (2,4- DIENOYL COENZYME A REDUCTASE)		reductase	1006
1646	80232038 (7625, 7626)	Novel Protein sim. GBank gi 1186971 sp P18775 DMSA_ECOLI - ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN A PRECURSOR (DMSO REDUCTASE)		reductase	1004, 1006, 1009, 1010, 1022, 1024, 1031
1647	18594306 (9955, 9956)	Novel Protein sim. GBank gi 128515 sp P19573 NOSZ_PSEST - NITROUS- OXIDE REDUCTASE PRECURSOR (N2)OR		reductase	1024
1648	78768523 (16791, 16792)	Novel Protein sim. GBank gi 1346662 sp P09152 NARG_ECOLI - RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN		reductase	1017
1649	65490548 (11399, 11400)	Novel Protein sim. GBank gi 1351892 sp P48360 ADRO_YEAST - NADPH:ADRENODOXIN OXIDOREDUCTASE HOMOLOG PRECURSOR (ADRENODOXIN REDUCTASE HOMOLOG)		reductase	1001, 1006, 1009, 1010, 1012, 1022, 1023, 1034
1650	51971443 (1801, 1802)	Novel Protein sim. GBank gi 1352934 sp P47169 YJ9F_YEAST - HYPOTHETICAL 161.2 KD PROTEIN IN NMD5- HOM6 INTERGENIC REGION		reductase	1040

1651	80240231 (3, 4)	Novel Protein sim. GBank gil136691 sp P00126 UCRH_BOVIN - UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KD PROTEIN (MITOCHONDRIAL HINGE PROTEIN) (CYTOCHROME C1, NONHEMЕ 11 KD PROTEIN) (COMPLEX III SUBUNIT VIII)		reductase	1004, 1038
1652	27977546 (16227, 16228)	Novel Protein sim. GBank gil1651696 dbj BAA16624 - (D90899) peptide methionine sulfoxide reductase [Synechocystis sp.]		reductase	1006
1653	27977581 (21163, 21164)	Novel Protein sim. GBank gil1651696 dbj BAA16624 - (D90899) peptide methionine sulfoxide reductase [Synechocystis sp.]		reductase	1006
1654	19748869 (1005, 1006)	Novel Protein sim. GBank gil166611 emb CAB05057 - (Z82098) hypothetical protein Rv3549c [Mycobacterium tuberculosis]		reductase	1039
1655	78688051 (2755, 2756)	Novel Protein sim. GBank gil171399 (M58012) - ribonucleotide reductase DNA damage-inducible regulatory subunit [Saccharomyces cerevisiae]		reductase	1008
1656	79574869 (18729, 18730)	Novel Protein sim. GBank gil1717781 sp P29509 TRXB_YEAST - THIOREDOXIN REDUCTASE		reductase	1026, 1031
1657	20467247 (4453, 4454)	Novel Protein sim. GBank gil1723442 sp Q10258 YD2A_SCHPO - HYPOTHETICAL 69.0 KD PROTEIN C56FF8.10 IN CHROMOSOME I		reductase	1010
1658	23294536 (19841, 19842)	Novel Protein sim. GBank gil1723913 sp P53128 YGM5_YEAST - HYPOTHETICAL 68.5 KD PROTEIN IN SCS3-SUP44 INTERGENIC REGION		reductase	1044
1659	20294915 (8197, 8198)	Novel Protein sim. GBank gil1773103 (U82664) - similar to S_cerevisiae_l-pg20p [Escherichia coli]		reductase	1022

1660	80504192 (14141, 14142)	Novel Protein sim. GBank gi 1806154 emb CAB06451 - (Z84395) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		reductase	1003, 1006, 1008, 1012, 1017, 1022, 1023, 1024, 1027, 1031, 1034, 1044
1661	20465254 (4775, 4776)	Novel Protein sim. GBank gi 2078004 emb CAB08451 - (Z95207) gora [Mycobacterium tuberculosis]		reductase	1010, 1034
1662	79601001 (12325, 12326)	Novel Protein sim. GBank gi 2078004 emb CAB08451 - (Z95207) gora [Mycobacterium tuberculosis]		reductase	1022, 1023, 1026, 1029, 1036
1663	86691473 (2979, 2980)	Novel Protein sim. GBank gi 2105093 emb CAA71210 - (Y10124) alpha subunit of nitrate reductase [Thermus thermophilus]		reductase	1018
1664	79816887 (16019, 16020)	Novel Protein sim. GBank gi 2117199 emb CAB09020 - (Z95584) narG [Mycobacterium tuberculosis]		reductase	1016, 1029, 1034
1665	87458141 (22791, 22792)	Novel Protein sim. GBank gi 2117199 emb CAB09020 - (Z95584) narG [Mycobacterium tuberculosis]		reductase	1018, 1034
1666	78799757 (4023, 4024)	Novel Protein sim. GBank gi 2117201 emb CAB09006 - (Z95584) narJ [Mycobacterium tuberculosis]		reductase	1009
1667	56316218 (7639, 7640)	Novel Protein sim. GBank gi 2132110 pir S67182 - hypothetical protein YOR280C - yeast (Saccharomyces cerevisiae)		reductase	1026
1668	8901633 (7839, 7840)	Novel Protein sim. GBank gi 2132926 pir S67139 - probable membrane protein YQR246c - yeast (Saccharomyces cerevisiae)		reductase	1004
1669	95292749 (17735, 17736)	Novel Protein sim. GBank gi 2494779 sp Q47266 HMPA_ERWCH - FLAVOHEMOPROTEIN (HAEMOGLOBIN-LIKE PROTEIN) (FLAVOHEMOGLOBIN) (HEMOPROTEIN X)		reductase	1010, 1022, 1044

1670	79907207 (8887, 8888)	Novel Protein sim. GBank gil24956281 sp P55757 YOHLYHLSERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BIOA 5'REGION		reductase	1003, 1008, 1017, 1023, 1039
1671	359992073 (8873, 8874)	Novel Protein sim. GBank gil2498291 sp P72642 DAPB_SYNY3 - DIHYDRODIPICOLINATE REDUCTASE		reductase	1030
1672	36825635 (2213, 2214)	Novel Protein sim. GBank gil2499420 sp P77258 NEMA_ECOLI - N-ETHYLMALEIMIDE REDUCTASE (N-ETHYLMALEIMIDE REDUCING ENZYME)		reductase	1054
1673	656893774 (1225, 1226)	Novel Protein sim. GBank gil2507291 sp P35340 AHPF_ECOLI - ALKYL HYDROPEROXIDE REDUCTASE SUBUNIT F (ALKYL HYDROPEROXIDE REDUCTASE F52A PROTEIN)		reductase	1016
1674	28456387 (10495, 10496)	Novel Protein sim. GBank gil2565334 (AF026066) - sulfite reductase [Pseudomonas aeruginosa]		reductase	1003
1675	78378323 (19561, 19562)	Novel Protein sim. GBank gil2584874 (AF002661) - nitric oxide reductase [Ralstonia eutropha]		reductase	1008, 1026
1676	23302385 (22805, 22806)	Novel Protein sim. GBank gil2621389 (AE0000818) - GDP-D-mannose dehydratase [Methanobacterium thermoautotrophicum]		reductase	1026
1677	20446451 (13317, 13318)	Novel Protein sim. GBank gil269599 sp emb CAA15852 - (AL010186) fadH [Mycobacterium tuberculosis]		reductase	1024
1678	78746449 (7541, 7542)	Novel Protein sim. GBank gil2765041 sp CAA71086 - (Y09960) 2-enote reductase [Clostridium tyrobutyricum]		reductase	1017
1679	29024773 (20571, 20572)	Novel Protein sim. GBank gil2765041 sp CAA71086 - (Y09960) 2-enote reductase [Clostridium tyrobutyricum]		reductase	1017
1680	79829484 (11031, 11032)	Novel Protein sim. GBank gil282964 sp P74101 Y195_SYNY3 - HYPOTHETICAL 79.7 KD PROTEIN SLL1895		reductase	1003, 1017

1681	79186294 (19175, 19176)	Novel Protein sim. GBank gj[2982990 (AE000682)] - hypothetical protein [Aequifex aeolicus]		reductase	1003, 1024, 1026
1682	10343299 (22265, 22266)	Novel Protein sim. GBank gj[2983832 (AE000740)] - 5,10-methylenetetrahydrofolate reductase [Aequifex aeolicus]		reductase	1038
1683	79265417 (2341, 2342)	Novel Protein sim. GBank gj[2995313]emb CAA18341] - (AL022268) ribonucleotide reductase [Streptomyces coelicolor]		reductase	1013
1684	86472188 (3079, 3080)	Novel Protein sim. GBank gj[2995313]emb CAA18341] - (AL022268) ribonucleotide reductase [Streptomyces coelicolor]		reductase	1025, 1029
1685	79840520 (8489, 8490)	Novel Protein sim. GBank gj[3063883]emb CAA18568] - (AL022486) putative 3-oxoacyl-fatty-carrier protein] reductase [Mycobacterium leprae]		reductase	1008
1686	85899153 (11801, 11802)	Novel Protein sim. GBank gj[3282229]gb AAC24957.1] - (U68722) BCLHH [Botryotinia fuckeliana]		reductase	1025, 1029
1687	11090590 (6657, 6658)	Novel Protein sim. GBank gj[3329297 (AE001355)] - Ribonucleoside Reductase, Large Chain [Chlamydia trachomatis]		reductase	1006
1688	78461591 (14279, 14220)	Novel Protein sim. GBank gj[3402836]emb CAA76082] - (Y16136) 2-enoyl reductase [Moorella thermoacetica]		reductase	1044
1689	5288460 (13957, 13958)	Novel Protein sim. GBank gj[3482915 (AC003970)] - Similar to 12-oxophytodienoate reductase, gj[2765083 and old-yellow-enzyme homolog, gj[2232254 [Arabidopsis thaliana]		reductase	1058
1690	20451162 (10513, 10514)	Novel Protein sim. GBank gj[3560010]emb CAA20632] - (AL031515) putative nitrate reductase alpha chain [Streptomyces coelicolor]		reductase	1024

1691	65997973 (12235) 12236)	Novel Protein sim. GBank gi 3560010 emb CAA20632 - (AL031515) putative nitrate reductase alpha chain [Streptomyces coelicolor]		reductase	1024
1692	20463334 (4927, 4928)	Novel Protein sim. GBank gi 3805970 emb CAA06231 - (AJ004933) periplasmic nitrate reductase, large subunit [Rhodopseudomonas sp.]		reductase	1010
1693	25137368 (12555) 12556)	Novel Protein sim. GBank gi 3913677 spi Q12473 FRE6 YEAST - FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 6 PRECURSOR		reductase	1008
1694	11129482 (14641, 14642)	Novel Protein sim. GBank gi 3913677 spi Q12473 FRE6 YEAST - FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 6 PRECURSOR		reductase	1006
1695	65678005 (6505, 6506)	Novel Protein sim. GBank gi 4007748 emb CAA07451 - (AJ007313) thioredoxin reductase (NADPH) [Streptomyces coelicolor]		reductase	1010
1696	39514968 (15049, 15050)	Novel Protein sim. GBank gi 4096117 gb AAC99805.1 - (U24271) tetra hydrofolate reductase [Saccharomyces cerevisiae]		reductase	1010
1697	80218341 (21431, 21432)	Novel Protein sim. GBank gi 418423 spi P32614 YEF7 YEAST - HYPOTHETICAL 50.8 KD PROTEIN IN PAU2- GLY1 INTERGENIC REGION		reductase	1006, 1016, 1022, 1025
1698	94678667 (17611, 17612)	Novel Protein sim. GBank gi 42121 emb CAA32416 - (X14202) NADH- dependent nitrite reductase (nirB) (AA 1-839) [Escherichia coli]		reductase	1006
1699	47652339 (19681, 19682)	Novel Protein sim. GBank gi 4836502 gb AAD30468.1 - (AF124929) putative aldo/keto reductase family 2 enzyme [Streptomyces clavuligerus]		reductase	1029, 1053

1700	80473319 (18493, 18494)	Novel Protein sim. GBank gi 520501 (U10274) - thioredoxin reductase [Saccharomyces cerevisiae]		reductase	1003, 1004, 1006, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1027, 1031, 1034, 1038, 1039, 1044, 1054
1701	20748619 (12253, 12254)	Novel Protein sim. GBank gi 5437841 sp P26427 AHPG_ECOLI-ALKYL HYDROPEROXIDE REDUCTASE C22 PROTEIN (SCRP-23) (SULFATE STARVATION-INDUCED PROTEIN 8) (SS18)		reductase	1006
1702	20465720 (15693, 15694)	Novel Protein sim. GBank gi 5457308 emb CAB46979.1 - (AJ243482) CSR14 protein [Enterococcus faecalis]		reductase	1010
1703	20288906 (22835, 22836)	Novel Protein sim. GBank gi 5459219 emb CAB48892.1 - (AL0968837) hypothetical protein [Streptomyces coelicolor A3(2)]		reductase	1034
1704	20293676 (5667, 5668)	Novel Protein sim. GBank gi 546494 bbs 147626 - (S70056) zeta-crystallin, nicotinamide adenine dinucleotide phosphate:quinone reductase [mice, liver, Peptide, 330 aa] [Mus sp.]		reductase	1034
1705	79249850 (2015, 2016)	Novel Protein sim. GBank gi 5497581 sp P36033 FRE2_YEAST - FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 2 PRECURSOR		reductase	1017, 1023, 1026, 1039
1706	47657054 (15829, 15830)	Novel Protein sim. GBank gi 5525066 emb CAB50882.1 - (AL096884) putative enoyl-(acyl-carrier-protein) reductase [Streptomyces coelicolor A3(2)]		reductase	1022, 1029

1707 3158)	56924278 (3157, Novel Protein sim. GBank gi 5855562 sp Q06458 NIRB_KLEPN - NITRITE REDUCTASE (NADPH) LARGE SUBUNIT		reductase	1003
1708 (22108)	78728722 (22107, Novel Protein sim. GBank gi 730074 sp P39692 MT10_YEAST - SULFITE REDUCTASE (NADPH) FLAVOPROTEIN COMPONENT		reductase	1022, 1026
1709 22202)	20297592 (22201, Novel Protein sim. GBank gi 732352 sp P39605 YWCG_BACSU - HYPOTHETICAL 28.3 KD PROTEIN IN QO(XD- VPR INTERGENIC REGION		reductase	1034
1710 4976)	27364802 (4975, Novel Protein sim. GBank gi 96217 pir S10792 - nirD protein - Escherichia coli		reductase	1009
1711 5550)	80497318 (5549, Novel Protein sim. GBank gi 4582468 gb AAD24852.1 AC0070707 - (AC007071) putative 40S ribosomal protein; contains C-terminal domain [Arabidopsis thaliana]		reductase	1012
1712 8604)	78905077 (8603, Novel Protein sim. GBank gi 4582468 gb AAD24852.1 AC0070707 - (AC007071) putative 40S ribosomal protein; contains C-terminal domain [Arabidopsis thaliana]	Contains protein domain (PF00013) - KH domain	ribosomalprot	1003, 1008, 1017
1713 19280)	37016632 (19279, Novel Protein sim. GBank gi 132869 sp P11255 RL2_YERRPS - 50S RIBOSOMAL PROTEIN L2	Contains protein domain (PF00181) - Ribosomal Proteins L2	ribosomalprot	1025
1714 19793)	86473978 (19793, Novel Protein sim. GBank gi 4582468 gb AAD24852.1 AC0070707 - (AC007071) putative 40S ribosomal protein; contains C-terminal domain [Arabidopsis thaliana]	Contains protein domain (PF00189) - Ribosomal protein S3, C-terminal domain.	ribosomalprot	1054
1715 13140)	11101120 (13139, Novel Protein sim. GBank gi 1173216 sp P44385 RS19_HAEIN - 30S RIBOSOMAL PROTEIN S19	Contains protein domain (PF00203) - Ribosomal protein S19	ribosomalprot	1006
1716 12796)	65645736 (12795, Novel Protein sim. GBank gi 132854 sp P02387 RL2_ECOLI - 50S RIBOSOMAL PROTEIN L2	Contains protein domain (PF00203) - Ribosomal protein S19	ribosomalprot	1041
1717 4060)	47651998 (4059, Novel Protein sim. GBank gi 3122702 sp Q06115 RL22_MYCSM - 50S RIBOSOMAL PROTEIN L22	Contains protein domain (PF00237) - Ribosomal protein L22	ribosomalprot	1029

1718	37026724 (2257, 2258)	Novel Protein sim. GBank gil122713 sp O32986 RL22_MYCLE - 50S RIBOSOMAL PROTEIN L22	Contains protein domain (PF00237) - Ribosomal protein L22	ribosomalprot	1038
1719	79606269 (13597, 13598)	Novel Protein sim. GBank gil11729561 sp P46176 RL14_ACYKS - 50S RIBOSOMAL PROTEIN L14	Contains protein domain (PF00238) - Ribosomal protein L14	ribosomalprot	1022
1720	86473516 (615, 616)	Novel Protein sim. GBank gil2982289 (AF051229) - 60S ribosomal protein L17 [Picea mariana]	Contains protein domain (PF00238) - Ribosomal protein L14	ribosomalprot	1054
1721	80256964 (1781, 1782)	Novel Protein sim. GBank gil3122692 sp O32988 RL16_MYCLE - 50S RIBOSOMAL PROTEIN L16	Contains protein domain (PF00252) - Ribosomal protein L16	ribosomalprot	1006
1722	9894255 (3669, 3670)	Novel Protein sim. GBank gil3861192 emb CAA15092 - (AJ235272) 50S RIBOSOMAL PROTEIN L16 (rpIP) [Rickettsia prowazekii]	Contains protein domain (PF00252) - Ribosomal protein L16	ribosomalprot	1034
1723	37801612 (16899, 16900)	Novel Protein sim. GBank gil3777600 (AF095708) - 50S ribosomal protein L5 [Oryza sativa]	Contains protein domain (PF00281) - Ribosomal protein L5	ribosomalprot	1058
1724	20439217 (16943, 16944)	Novel Protein sim. GBank gil132954 sp P02386 RL3_ECOLI - 50S RIBOSOMAL PROTEIN L3	Contains protein domain (PF00297) - Ribosomal protein L3	ribosomalprot	1024
1725	85549924 (3799, 3800)	Novel Protein sim. GBank gil2500251 sp P96931 RL11_MYCTU - 50S RIBOSOMAL PROTEIN L11	Contains protein domain (PF00298) - Ribosomal protein L11	ribosomalprot	1025, 1029
1726	80055007 (5355, 5356)	Novel Protein sim. GBank gil1173023 sp P46789 RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00327) - Ribosomal protein L30p/L7e	ribosomalprot	1004, 1010, 1024, 1029, 1034, 1049,
1727	30174462 (20503, 20504)	Novel Protein sim. GBank gil132870 sp P0243 RL30_BACST - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00327) - Ribosomal protein L30p/L7e	ribosomalprot	1053, 1030
1728	11072079 (12609, 12610)	Novel Protein sim. GBank gil133964 sp P02357 RS5_BACST - 30S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	1034

1729	20458307 (2953, 2954)	Novel Protein sim. GBank gil1710791 sp Q10234 RT05_SCHPO - PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	1024
1730	79174828 (10275, 10276)	Novel Protein sim. GBank gil1363775 pir S59279 - Contains protein domain (PF00338) - Ribosomal protein S10p/S20e	ribosomalprot	1004, 1008, 1017, 1024, 1026, 1044	
1731	80243214 (17069, 17070)	Novel Protein sim. GBank gil1330361 sp P05738 RL9A_YEAST - 60S RIBOSOMAL PROTEIN L9 A (YL11) (RP25)	Contains protein domain (PF00347) - Ribosomal protein L6	ribosomalprot	1001, 1003, 1004, 1006, 1008, 1009, 1017, 1022, 1023, 1024, 1026, 1034, 1038, 1044
1732	78374701 (4735, 4736)	Novel Protein sim. GBank gil2058273 dbj BAA19798 - (D83527) YK426 [Oryza sativa]	Contains protein domain (PF00347) - Ribosomal protein L6	ribosomalprot	1003
1733	80059426 (11193, 11194)	Novel Protein sim. GBank gil1340281 sp P07842 RS9_BACST - 30S RIBOSOMAL PROTEIN S9 (BS10)	Contains protein domain (PF00380) - Ribosomal protein S9/S16	ribosomalprot	1010, 1024
1734	13088010 (21793, 21794)	Novel Protein sim. GBank gil285397 pir B43310 - ribosomal protein S9 homolog - Haemophilus somnus	Contains protein domain (PF00380) - Ribosomal protein S9/S16	ribosomalprot	1031
1735	14401735 (8259, 8260)	Novel Protein sim. GBank gil2500419 sp P56015 RS8_HELPY - 30S RIBOSOMAL PROTEIN S8	Contains protein domain (PF00410) - Ribosomal protein S8	ribosomalprot	1053
1736	80059396 (3371, 3372)	Novel Protein sim. GBank gil11731178 sp P44379 RS11_HAEIN - 30S RIBOSOMAL PROTEIN S11	Contains protein domain (PF00411) - Ribosomal protein S11	ribosomalprot	1024
1737	95002835 (3867, 3868)	Novel Protein sim. GBank gil2760184 dbj BAA24193.1 - (AB010082) ribosomal protein S11 [Bacillus halodurans]	Contains protein domain (PF00411) - Ribosomal protein S11	ribosomalprot	1022
1738	10081577 (945, 946)	Novel Protein sim. GBank gil1787938 (AE000260) - orf, hypothetical protein [Escherichia coli]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	ribosomalprot	1017

1739	80043804 (19703, 19704)	Novel Protein sim. GBank gil1870009 emb CAB06860 - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	ribosomalprot	1006, 1034
1740	20292795 (2937, 2938)	Novel Protein sim. GBank gil2495431 sp Q11063 Y08R MYCTU - HYPOTHETICAL TRANSCRIPTIONAL REGULATOR CY50_27	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	ribosomalprot	1034
1741	57531207 (16069, 16070)	Novel Protein sim. GBank gil73098 pir RPECR1 - tet repressor protein (Tn 1721) - Escherichia coli plasmid RP1	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	ribosomalprot	1004, 1010, 1012, 1024
1742	82051891 (17493, 17494)	Novel Protein sim. GBank gil3581853 emb CAA20809 - (AL031541) 50S ribosomal protein L20 [Streptomyces coelicolor]	Contains protein domain (PF00453) - Ribosomal protein L20	ribosomalprot	1004, 1006, 1010, 1016, 1022, 1023, 1024, 1031, 1034, 1039, 1044
1743	28970215 (10879, 10880)	Novel Protein sim. GBank gil2132069 pir S67038 - Contains protein domain (PF00572) - Ribosomal protein L13	Contains protein domain (PF00572) - Ribosomal protein L13	ribosomalprot	1044
1744	79192013 (17241, 17242)	Novel Protein sim. GBank gil2500256 sp Q53874 RL13_STRCO - 50S RIBOSOMAL PROTEIN L13	Contains protein domain (PF00572) - Ribosomal protein L13	ribosomalprot	1024, 1044
1745	85803561 (14581, 14582)	Novel Protein sim. GBank gil133870 sp P14128 RS1_PROSP - 30S RIBOSOMAL PROTEIN S1	Contains protein domain (PF00576) - S1 RNA binding domain	ribosomalprot	1038
1746	24140021 (10279, 10280)	Novel Protein sim. GBank gil42837 emb CAA23630 - (Y00342) reading frame S1 [Escherichia coli]	Contains protein domain (PF00575) - S1 RNA binding domain	ribosomalprot	1003
1747	80469610 (20189, 20190)	Novel Protein sim. GBank gil77107 pir R5BY16 - ribosomal protein L11.e.B, cytosolic - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00673) - ribosomal L5P. family C-terminus	ribosomalprot	1004, 1006, 1012, 1017, 1022, 1024, 1034, 1044
1748	20438082 (11181, 11182)	Novel Protein sim. GBank gil71060 pir R5BS1 - ribosomal protein L1 - Bacillus stearothermophilus	Contains protein domain (PF00687) - L1P family of ribosomal proteins	ribosomalprot	1024

1749	79199246 (19229, 19230)	Novel Protein sim. GBank gil132837 sp P04457 RL29_BACST - 50S RIBOSOMAL PROTEIN L29	Contains protein domain (PF00831) - Ribosomal L29 protein	ribosomalprot	1024
1750	30454170 (14957, 14958)	Novel Protein sim. GBank gil132840 sp P02429 RL29_ECOLI - 50S RIBOSOMAL PROTEIN L29	Contains protein domain (PF00831) - Ribosomal L29 protein	ribosomalprot	1008, 1012
1751	65494686 (20787, 20788)	Novel Protein sim. GBank gil3191984 emb CAA19383 - (AL023797) 30S ribosomal protein S16 [Streptomyces coelicolor]	Contains protein domain (PF00886) - Ribosomal protein S16	ribosomalprot	1023
1752	80578448 (56663, 56664)	Novel Protein sim. GBank gil1350986 sp P49397 RS3A_ORYSA - 40S RIBOSOMAL PROTEIN S3A (CYC07 PROTEIN)	Contains protein domain (PF01015) - Ribosomal S3Ae family	ribosomalprot	1034, 1044, 1058
1753	87453131 (4393, 4394)	Novel Protein sim. GBank gil3885884 (AF093630) - 60S ribosomal protein L21 [Oryza sativa]	Contains protein domain (PF01157) - Ribosomal protein L21e	ribosomalprot	1001, 1025
1754	80439624 (3401, 3402)	Novel Protein sim. GBank gil3298443 dbj BA31512 - (AB010880) chloroplast ribosomal protein L17 [Nicotiana tabacum]	Contains protein domain (PF01196) - Ribosomal protein L17	ribosomalprot	1038
1755	80440301 (7979, 7980)	Novel Protein sim. GBank gil730590 sp P22353 RM08_YEAST - MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L8 (YML8)	Contains protein domain (PF01196) - Ribosomal protein L17	ribosomalprot	1001, 1003, 1006, 1008, 1017, 1022, 1024, 1025, 1026, 1031, 1034, 1038, 1039
1756	80502966 (16955, 16956)	Novel Protein sim. GBank gil1173025 sp P44367 RL31_HAEIN - 50S RIBOSOMAL PROTEIN L31	Contains protein domain (PF01197) - Ribosomal protein L31	ribosomalprot	1012
1757	80249528 (17421, 17422)	Novel Protein sim. GBank gil4808366 emb CAB42780_1 - (AL049841) putative 50S ribosomal protein L31 [Streptomyces coelicolor]	Contains protein domain (PF01197) - Ribosomal protein L31	ribosomalprot	1003, 1006, 1023
1758	20370371 (5167, 5168)	Novel Protein sim. GBank gil2899551 sp P71710RS6_MYCTU - 30S RIBOSOMAL PROTEIN S6	Contains protein domain (PF01250) - Ribosomal protein S6	ribosomalprot	1004

1759	91254029 (9167, 9168)	Novel Protein sim. GBank gi 3377797 (AF075597) - Similar to 60S ribosome protein L19; coded for by A. thaliana cDNA T04719; coded for by A. thaliana cDNA H36046; coded for by A. thaliana cDNA T44067; coded for by A. thaliana cDNA T14056; coded for by A. thaliana cDNA R90691 [Ara...]	Contains protein domain (PF01280) - Ribosomal protein L19e	ribosomalprot	1009, 1010, 1022, 1054, 1058
1760	78789951 (7047, 7048)	Novel Protein sim. GBank gi 2065225 emb CAB08287 - (Z95117) hypothetical protein MLC1351.14c [Mycobacterium leprae]	Contains protein domain (PF01479) - S4 domain	ribosomalprot	1008
1761	82371629 (8421, 8422)	Novel Protein sim. GBank gi 4539113 emb CAB39834.1 - (AL049491) putative 30S ribosomal protein S4 [Mycobacterium leprae]	Contains protein domain (PF01479) - S4 domain	ribosomalprot	1001, 1009
1762	94653751 (8141, 8142)	Novel Protein sim. GBank gi 1574779 (U32811) - ribosomal protein L35 (rpL35) [Haemophilus influenzae Rd]	Contains protein domain (PF01632) - Ribosomal protein L35	ribosomalprot	1010
1763	71456390 (20407, 20408)	Novel Protein sim. GBank gi 3128228 (AC004077) - putative ribosomal protein L18A [Arabidopsis thaliana]	Contains protein domain (PF01775) - Ribosomal L18ae protein family	ribosomalprot	1008, 1034
1764	71329807 (15129, 15130)	Novel Protein sim. GBank gi 4506843 ref NP_000989.1 rrPL3 - ribosomal protein L37a	Contains protein domain (PF01780) - Ribosomal L37ae protein family	ribosomalprot	1026
1765	12861828 (22053, 22054)	Novel Protein sim. GBank gi 1077574 pir S52681 - hypothetical protein YDR116c - yeast (Saccharomyces cerevisiae)		ribosomalprot	1031
1766	10366252 (14797, 14798)	Novel Protein sim. GBank gi 1172623 sp P44402 PRMA_HAEIN - RIBOSOMAL PROTEIN L11 METHYLTRANSFERASE		ribosomalprot	1001
1767	90935098 (15395, 15396)	Novel Protein sim. GBank gi 1172991 sp P46778 RL21_HUMAN - 60S RIBOSOMAL PROTEIN L21		ribosomalprot	1044
1768	79828082 (2267, 2268)	Novel Protein sim. GBank gi 1173000 sp P45581 RL25_HAEIN - 50S RIBOSOMAL PROTEIN L25		ribosomalprot	1039

1769	80217807 (3789, 3790)	Novel Protein sim. GBank gi 1176558 sp P46734 YM73_YEAST - PUTATIVE 40S RIBOSOMAL PROTEIN IN RRP5-END1 INTERGENIC REGION		ribosomalprot	1004, 1006, 1008, 1010, 1017, 1022, 1023, 1024, 1026, 1030, 1034, 1039
1770	13000915 (9625, 9626)	Novel Protein sim. GBank gi 132815 sp P02425 RL24_ECOLI - 50S RIBOSOMAL PROTEIN L24		ribosomalprot	1031
1771	79161541 (16833, 16834)	Novel Protein sim. GBank gi 133127 sp P28773 RT17_YEAST - MITOCHONDRIAL 40S RIBOSOMAL PROTEIN MRP17		ribosomalprot	1024
1772	65656001 (20001, 20002)	Novel Protein sim. GBank gi 1350717 sp P47833 RL30_CHICK - 60S RIBOSOMAL PROTEIN L30		ribosomalprot	1054
1773	47656185 (7811, 7812)	Novel Protein sim. GBank gi 1352914 sp P47150 J83_YEAST - PUTATIVE 40S RIBOSOMAL PROTEIN YJR113C		ribosomalprot	1009, 1029
1774	18357314 (12081, 12082)	Novel Protein sim. GBank gi 1550814 emb CA-A69256.1 - (Y07959) 60S acidic ribosomal protein P0 [Zea mays]		ribosomalprot	1024
1775	78470813 (17249, 17250)	Novel Protein sim. GBank gi 165303 dbj BAA17950 - (D90910) hypothetical protein [Synechocystis sp.]		ribosomalprot	1053
1776	25260960 (10251, 10252)	Novel Protein sim. GBank gi 1666144 emb CA803779 - (Z81451) obg [Mycobacterium tuberculosis]		ribosomalprot	1026
1777	79569281 (2061, 2062)	Novel Protein sim. GBank gi 1710521 sp P50889 RL24_HORVU - 60S RIBOSOMAL PROTEIN L24		ribosomalprot	1034
1778	27363848 (9917, 9918)	Novel Protein sim. GBank gi 1710546 sp P52866 RL36_DAUCA - 60S RIBOSOMAL PROTEIN L36		ribosomalprot	1017

1779	77869739 (9759, 9760)	Novel Protein sim. GBank gi 1723727 sp P53295 YG3Y YEAST - HYPOTHETICAL 41.0 KD PROTEIN IN YIP- CBP4 INTERGENIC REGION	ribosomalprot	1019, 1041
1780	97977933 (3641, 3642)	Novel Protein sim. GBank gi 1783294 dbj BAA13501 - (D87916) Obg [Streptomyces griseus]	ribosomalprot	1001, 1006, 1012, 1019, 1022, 1027, 1034
1781	20210896 (5095, 5096)	Novel Protein sim. GBank gi 1877389 emb CA07108 - (Z92772) rpJ [Mycobacterium tuberculosis]	ribosomalprot	1010
1782	86665034 (15917, 15918)	Novel Protein sim. GBank gi 2052129 emb CA08155 - (Z94752) rnmJ [Mycobacterium tuberculosis]	ribosomalprot	1019, 1030
1783	79580821 (5427, 5428)	Novel Protein sim. GBank gi 2260035 pf 1405340A - protein 40kD [Mus musculus]	ribosomalprot	1016
1784	79828481 (14963, 14964)	Novel Protein sim. GBank gi 250028 sp P56029 RL1_HELPPY - 50S RIBOSOMAL PROTEIN L1	ribosomalprot	1003
1785	10326187 (7815, 7816)	Novel Protein sim. GBank gi 2507509 sp P37346 YCFH_ECOLI - HYPOTHETICAL 29.8 KD PROTEIN IN HOLB- PTSG INTERGENIC REGION	ribosomalprot	1003
1786	17893133 (2721, 2722)	Novel Protein sim. GBank gi 3128228 (AC004077) - putative ribosomal protein L18A [Arabidopsis thaliana]	ribosomalprot	1008
1787	29347594 (12583, 12584)	Novel Protein sim. GBank gi 3128228 (AC004077) - putative ribosomal protein L18A [Arabidopsis thaliana]	ribosomalprot	1022
1788	66397081 (12807, 12808)	Novel Protein sim. GBank gi 3319759 emb CAA19925 - (AL031035) ribosome recycling factor [Streptomyces coelicolor]	ribosomalprot	1008
1789	86673895 (15761, 15762)	Novel Protein sim. GBank gi 3608479 (AF088912) - ribosomal protein L15 [Petunia x hybrida]	ribosomalprot	1007

1790	30258207 (5149, 5150)	Novel Protein sim. GBank gi 3914989 sp O22424 RS4_MAIZE - 40S RIBOSOMAL PROTEIN S4		ribosomal prot	1044
1791	37802328 (11141, 11142)	Novel Protein sim. GBank gi 4506719 ref NP_002943.1 pRPS2 - ribosomal protein S2		ribosomal prot	1058
1792	11129136 (18285, 18286)	Novel Protein sim. GBank gi 4588906 gb ADD26256.1 AF118144 - (AF118149) ribosomal protein S7 [Secale cereale] A3(2)]		ribosomal prot	1022
1793	80090625 (19759, 19760)	Novel Protein sim. GBank gi 5688891 emb CAB52054.1 - (AL109732) 30S ribosomal protein S1 [Streptomyces coelicolor A3(2)]		ribosomal prot	1004, 1008, 1026
1794	78383456 (8297, 8298)	Novel Protein sim. GBank gi 5856870 sp P05735 RL19_YEAST - 60S RIBOSOMAL PROTEIN L19 (L23) (YL14) (RP33) (RP15L)		ribosomal prot	1008
1795	23282406 (16267, 16268)	Novel Protein sim. GBank gi 5856876 sp Q07761 RL2B_TOBAC - 60S RIBOSOMAL PROTEIN L23A (L25)		ribosomal prot	1044
1796	57292746 (18055, 18056)	Novel Protein sim. GBank gi 5856898 sp P36531 RM36_YEAST - MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L36 PRECURSOR (YML36)		ribosomal prot	1004, 1006, 1010, 1022, 1034
1797	8757981 (16007, 16008)	Novel Protein sim. GBank gi 78837 pir S04775 - rrmK protein - Escherichia coli		ribosomal prot	1004
1798	28480961 (16645, 16646)			ribosomal prot	1008
1799	80028308 (15985, 15986)	Novel Protein sim. GBank gi 2500789 sp Q50712 Y09F_MYCTU - PUTATIVE RNA POLYMERASE SIGMA FACTOR CY78_15	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	mapolymerase	1006, 1044
1800	20429859 (19955, 19956)	Novel Protein sim. GBank gi 419697 pir JN0443 - transcription initiation factor sigma homolog hrdB - Streptomyces aureofaciens	Contains protein domain (PF00140) - Sigma-70 factor	mapolymerase	1010

1801	14206194 (4171, 4172)	Novel Protein sim. GBank g 482053 pir S40575 - hypothetical protein - Escherichia coli	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	mapolymerase	1034
1802	30007819 (6689, 6690)	Novel Protein sim. GBank g 282943 sp P52293 MA2_MOUSE - IMPORTIN ALPHA-2 SUBUNIT (KARYOPHERIN ALPHA-2 SUBUNIT) (SRP1-ALPHA) (RAG COHOR PROTEIN 1) (PENDULIN) (PORE TARGETING COMPLEX 58 KD SUBUNIT) (PTAC58) (IMPORTIN ALPHA P1)	Contains protein domain (PF00514) - Armadillo/beta-catenin-like repeats	mapolymerase	1022
1803	17929579 (13007, 13008)	Novel Protein sim. GBank g 1432083 (U60981) - homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana]	Contains protein domain (PF01466) - Skp1 family	mapolymerase	1017, 1044
1804	95287742 (4021, 4022)	Novel Protein sim. GBank g 1173155 sp P43739 RPOC_HAEIN - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT)		mapolymerase	1012
1805	20438830 (20637, 20638)	Novel Protein sim. GBank g 1173155 sp P43739 RPOC_HAEIN - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT)		mapolymerase	1024
1806	80258475 (15711, 15712)	Novel Protein sim. GBank g 1173288 sp P38106 RSEA_ECOLI - SIGMA-E FACTOR NEGATIVE REGULATORY PROTEIN		mapolymerase	1004
1807	79574446 (2171, 2172)	Novel Protein sim. GBank g 133294 sp P06223 RPS4_KLEPN - RNA POLYMERASE SIGMA-54 FACTOR		mapolymerase	1031
1808	20609900 (1151, 1152)	Novel Protein sim. GBank g 133321 sp P17531 RPSD_MYXXA - RNA POLYMERASE SIGMA FACTOR RPOD (SIGMA-80)		mapolymerase	1004

1809	78772569 (6967, 6968)	Novel Protein sim. GBank gi 133356 sp P04051 RPC1_YEAST - DNA- DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (C160)		rnapolymerase	1023
1810	35110282 (1003, 1004)	Novel Protein sim. GBank gi 133418 sp P19175 RPOB_PSEPU - DNA- DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnapolymerase	1001, 1029
1811	27975439 (1509, 1510)	Novel Protein sim. GBank gi 133418 sp P19175 RPOB_PSEPU - DNA- DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnapolymerase	1053
1812	55763524 (16781, 16782)	Novel Protein sim. GBank gi 133418 sp P19175 RPOB_PSEPU - DNA- DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnapolymerase	1023
1813	96663942 (18089, 18090)	Novel Protein sim. GBank gi 133418 sp P19175 RPOB_PSEPU - DNA- DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnapolymerase	1039
1814	25140324 (11, 12)	Novel Protein sim. GBank gi 133424 sp P06271 RPOB_TOBAC - DNA- DIRECTED RNA POLYMERASE BETA CHAIN		rnapolymerase	1026
1815	16411357 (22613, 22614)	Novel Protein sim. GBank gi 133451 sp P06221 RPOL_BRS6P6 - DNA- DIRECTED RNA POLYMERASE		rnapolymerase	1054
1816	20619650 (13135, 13136)	Novel Protein sim. GBank gi 133457 sp P13433 RPOM_YEAST - DNA- DIRECTED RNA POLYMERASE MITOCHONDRIAL PRECURSOR		rnapolymerase	1006

1817	37004231 (22497, 22498)	Novel Protein sim. GBank gil1350849 spi P47768 RPOB_STAAU - DNA-DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnpolymerase	1029
1818	78381427 (22775, 22776)	Novel Protein sim. GBank gil1350849 spi P47768 RPOB_STAAU - DNA-DIRECTED RNA POLYMERASE BETA CHAIN (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT)		rnpolymerase	1008
1819	14994678 (6091, 6092)	Novel Protein sim. GBank gil147730 (M38305) - RNA polymerase (rpoC) [Escherichia coli]		rnpolymerase	1024
1820	29207782 (20633, 20634)	Novel Protein sim. GBank gil170956 (M96690) - A49 [Saccharomyces cerevisiae]		rnpolymerase	1012, 1024, 1044
1821	11305638 (3915, 3916)	Novel Protein sim. GBank gil170984 (M20315) - acidic Protein 2 [Saccharomyces cerevisiae]		rnpolymerase	1022
1822	30254064 (3393, 3394)	Novel Protein sim. GBank gil1710654 spi Q01080 RPA3 YEAST - DNA-DIRECTED RNA POLYMERASE I 49 KD POLYPEPTIDE (A49)		rnpolymerase	1026
1823	11607684 (2197, 2198)	Novel Protein sim. GBank gil172215 (J03530) - RNA polymerase A(I) large subunit [Saccharomyces cerevisiae]		rnpolymerase	1024
1824	23226319 (21891, 21892)	Novel Protein sim. GBank gil172215 (J03530) - RNA polymerase A(I) large subunit [Saccharomyces cerevisiae]		rnpolymerase	1044
1825	79262544 (13241, 13242)	Novel Protein sim. GBank gil2244822 emb [CAB10245.1] - (Z97336) RNA polymerase II fifth largest subunit like protein [Arabidopsis thaliana]		rnpolymerase	1003, 1017, 1039
1826	178802319 (7889, 7890)	Novel Protein sim. GBank gil3021676 dhb [BA2A25358] - (D86033) RNA polymerase sigma-70 factor [Pseudomonas fluorescens]		rnpolymerase	1039
1827	78382425 (17987, 17988)	Novel Protein sim. GBank gil3372230 (AF017074) - RNA polymerase I, II and III 16.5 kDa subunit [Arabidopsis thaliana]		rnpolymerase	1017, 1023

1828	87456505 (6035, 6036)	Novel Protein sim. GBank gj 3860708 emb CAA14609 - (AJ235270) DNA-DIRECTED RNA POLYMERASE BETA PRIME CHAIN [tpaC] [Rickettsia prowazekii]		rnapolymerase	1025
1829	47654603 (7053, 7054)	Novel Protein sim. GBank gj 3879980 emb CAB07490 - (Z92972) similar to Reverse transcriptase (RNA-dependent DNA polymerase) [Caenorhabditis elegans]		rnapolymerase	1026
1830	35105000 (4525, 4526)	Novel Protein sim. GBank gj 3914806 sp Q08429 RP54_VIBAN - RNA POLYMERASE SIGMA-54 FACTOR		rnapolymerase	1029
1831	11104094 (16841, 16842)	Novel Protein sim. GBank gj 3915852 sp P47769 RPC_MYCTU - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT)		rnapolymerase	1022
1832	21428191 (13579, 13580)	Novel Protein sim. GBank gj 401119 sp Q02821 IMA1_YEAST - IMPORTIN ALPHA SUBUNIT (KARYOPHERIN ALPHA SUBUNIT) (SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN)		rnapolymerase	1022
1833	86474962 (417, 418)	Novel Protein sim. GBank gj 409789 U00006 - DNA-directed RNA polymerase, beta'-subunit [Escherichia coli]		rnapolymerase	1038
1834	80503184 (6149, 6150)	Novel Protein sim. GBank gj 409789 U00006 - DNA-directed RNA polymerase, beta'-subunit [Escherichia coli]		rnapolymerase	1012
1835	79864126 (20625, 20626)	Novel Protein sim. GBank gj 4129212 AF095748 - principal sigma factor [Burkholderia cepacia]		rnapolymerase	1012
1836	25151061 (8255, 8256)	Novel Protein sim. GBank gj 417698 sp P32910 RPC6_YEAST - DNA-DIRECTED RNA POLYMERASE III 36 KD POLYPEPTIDE (C34)		rnapolymerase	1034

1837	19882587 (11429, 11430)	Novel Protein sim. GBank gi 4176981 sp P32910 RPC6_YEAST - DNA-DIRECTED RNA POLYMERASE II / 36 KD POLYPEPTIDE (C34)		rnopolymerase	1022
1838	11778342 (6865, 6866)	Novel Protein sim. GBank gi 4196599 pir JN0445 - transcription initiation factor sigma homolog hrdE - Streptomyces aluteofaciens		rnopolymerase	1006
1839	20728283 (6687, 6688)	Novel Protein sim. GBank gi 4467280 emb CAB37604 - (AL035570) putative dna-directed rna polymerase iii largest subunit [Schizosaccharomyces pombe]		rnopolymerase	1006
1840	80079561 (8461, 8462)	Novel Protein sim. GBank gi 4512395 dbj BAA75263_1 - (AB017508) rpoB homologus (identity of 87% to B. subtilis) [Bacillus halodurans]		rnopolymerase	1034
1841	86658876 (13929, 13930)	Novel Protein sim. GBank gi 4826936 ref NP_05026..1 pPOLR - polymerase (RNA) mitochondrial (DNA directed)		rnopolymerase	1008
1842	10121246 (11651, 11652)	Novel Protein sim. GBank gi 4982169 gb AAD36665..1 AE00180 - (AE001804) RNA polymerase sigma-E factor [Thermotoga maritima]		rnopolymerase	1023
1843	80500253 (17513, 17514)	Novel Protein sim. GBank gi 5155685 (U12205) - RNA-polymerase beta subunit [Mycobacterium tuberculosis]		rnopolymerase	1012, 1024
1844	47656628 (4031, 4032)	Novel Protein sim. GBank gi 559513 pir A47718 - reverse transcriptase, pol-like - chicken (fragment)		rnopolymerase	1029
1845	30811804 (8473, 8474)	Novel Protein sim. GBank gi 539907 pir A45983 - lactose-binding lectin Mac-2 - mouse		rnopolymerase	1026
1846	79263338 (17107, 17108)	Novel Protein sim. GBank gi 5531433 emb CAB50938..1 - (AL096849) putative RNA polymerase sigma factor [Streptomyces coelicolor A3(2)]		rnopolymerase	1039

1847	91254002 (4695, 4696)	Novel Protein sim. GBank gi 5712306 gb AAD47422.1 AF11185 - (AF11185) RNA polymerase alpha subunit [Agrobacterium tumefaciens]		mapolymerase	1010
1848	86464908 (11233, 11234)	Novel Protein sim. GBank gi 57403 emb CA451302 - (X72787) RNA-polymerase beta subunit [Streptomyces griseus]		mapolymerase	1008, 1009, 1024, 1029, 1031, 1044
1849	78732427 (19557, 19558)	Novel Protein sim. GBank gi 539298 pir S37958 - myosin heavy chain homolog YKL129c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00018) - SH3 domain	struct	1003, 1008, 1009, 1022, 1026
1850	78955619 (12571, 12572)	Novel Protein sim. GBank gi 113305 sp P10993 ACT2_TETPY - ACTIN, CYTOPLASMIC (ACTIN_MICRONUCLEAR)	Contains protein domain (PF00022) - Actin	struct	1030, 1058
1851	15035096 (14085, 14086)	Novel Protein sim. GBank gi 1168330 sp P45888 ARP2_DROME - ACTIN-LIKE PROTEIN 2 (ACTIN-LIKE PROTEIN 14D)	Contains protein domain (PF00022) - Actin	struct	1054
1852	30454535 (17969, 17970)	Novel Protein sim. GBank gi 1703154 sp P53502 ACT_FUCDI - ACTIN	Contains protein domain (PF00022) - Actin	struct	1030
1853	80084619 (9349, 9350)	Novel Protein sim. GBank gi 1008665 pir S22860 - calmodulin 2 (clone lambda DASH) - rice	Contains protein domain (PF00036) - EF hand	struct	1009, 1034, 1058
1854	78081281 (6145, 6146)	Novel Protein sim. GBank gi 115511 sp P13556 CALM_HORVU - CALMODULIN	Contains protein domain (PF00036) - EF hand	struct	1000
1855	11305935 (19191, 19192)	Novel Protein sim. GBank gi 3183044 sp Q63279 K1CS_RAT - KERATIN, TYPE I CYTOSKELETAL 19 (CYTOKERATIN 19) (K19) (CK 19)	Contains protein domain (PF00038) - Intermediate filament proteins	struct	1022
1856	27933240 (8233, 8234)	Novel Protein sim. GBank gi 4504383 ref NP_001519.1 pHGFA - HGF activator	Contains protein domain (PF00040) - Fibronectin type II domain	struct	1024
1857	80049617 (6233, 6234)	Novel Protein sim. GBank gi 3243131 (AF045777) - titin [Drosophila melanogaster]	Contains protein domain (PF00047) - Immunoglobulin domain	struct	1016, 1022

1858	57056391 (2901, 2902)	Novel Protein sim. GBank gil4099555 (U88578) - immunoglobulin superfamily member [Drosophila melanogaster]	Contains protein domain (PF00047) - immunoglobulin domain	struct	1009
1859	11801219 (2761, 2762)	Novel Protein sim. GBank gil4507721 refNP_003310.1 pTTN - titin	Contains protein domain (PF00047) - immunoglobulin domain	struct	1022
1860	80058910 (11305, 11306)	Novel Protein sim. GBank gil4507721 refNP_003310.1 pTTN - titin	Contains protein domain (PF00047) - immunoglobulin domain	struct	1022, 1025, 1030
1861	79572425 (14741, 14742)	Novel Protein sim. GBank gil4589628 db BA76836.1 - (AB023209) KIAA0992 protein [Homo sapiens]	Contains protein domain (PF00047) - immunoglobulin domain	struct	1001, 1051, 1054
1862	80221181 (12019, 12020)	Novel Protein sim. GBank gil127736 sp P19524 MYS2_YEAST - MYOSIN-2 ISOFORM	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	1006, 1008, 1010, 1022, 1023, 1034, 1038, 1044
1863	79837753 (20583, 20584)	Novel Protein sim. GBank gil127736 sp P19524 MYS2_YEAST - MYOSIN-2 ISOFORM	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	1008
1864	296911769 (15911, 15912)	Novel Protein sim. GBank gil417335 sp P32492 MYS4_YEAST - MYOSIN-4 ISOFORM	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	1009
1865	29688833 (7837, 7838)	Novel Protein sim. GBank gil4885026 gb AAD31926.1 AF14773 - (AF147738) myosin VIII ZMM3 [Zea mays]	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	1008
1866	29005698 (17363, 17364)	Novel Protein sim. GBank gil4885026 gb AAD31926.1 AF14773 - (AF147738) myosin VIII ZMM3 [Zea mays]	Contains protein domain (PF00063) - Myosin head (motor domain)	struct	1017
1867	29002174 (19913, 19914)	Novel Protein sim. GBank gil113947 sp P07150 ANX1_RAT - ANNEXIN I (LIPOCORTIN I) (CALPACTIN II) (CHROMOBINDIN 9) (P35) (PHOSPHOLIPASE A2 INHIBITORY PROTEIN)	Contains protein domain (PF00191) - Annexin	struct	1039
1868	79951559 (15, 16)	Novel Protein sim. GBank gil1723958 sp P53086 YGW6_YEAST - PUTATIVE KINESIN-LIKE PROTEIN YGL216W	Contains protein domain (PF00225) - Kinesin motor domain	struct	1001

1869	80219770 (2867, 2868)	Novel Protein sim. GBank g!1723958 sp!P53086 YGW6_YEAST - PUTATIVE KINESIN-LIKE PROTEIN YGL216W	Contains protein domain (PF002225) - Kinesin motor domain	struct	1001, 1006, 1010, 1022, 1024, 1034
1870	37802586 (965, 966)	Novel Protein sim. GBank g!3452553 dbj BAA32473 - (AB001347) brain beta 3 spectrin [Rattus norvegicus]	Contains protein domain (PF00435) - Spectrin repeat	struct	1058
1871	11116197 (3759, 3760)	Novel Protein sim. GBank g!1480339 emb CAA67770 - (X99333) scavenger receptor protein [Sus scrofa]	Contains protein domain (PF00530) - Scavenger receptor cysteine-rich domain	struct	1053
1872	33265035 (22305, 22306)	Novel Protein sim. GBank g!2135121 pif A49674 - flightless-I homolog - human (fragment)	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	1029
1873	17870815 (18033, 18034)	Novel Protein sim. GBank g!4567754 3 ref NP_001384_1 PECM2 - extracellular matrix protein 2	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	1003
1874	80027296 (8589, 8590)	Novel Protein sim. GBank g!869200(U29380) - similar to adenylylate cyclase [Caenorhabditis elegans]	Contains protein domain (PF00560) - Leucine Rich Repeat	struct	1006, 1010
1875	57303483 (17729, 17730)		Contains protein domain (PF00612) - IQ calmodulin- binding motif	struct	1004
1876	30259520 (6517, 6518)	Novel Protein sim. GBank g!3991173 sp!P31725 S109_MOUSE - CALGRANULIN B (MIGRATION INHIBITORY FACTOR-RELATED PROTEIN 14) (MRP-14) (P14) (LEUKOCYTE L1 COMPLEX HEAVY CHAIN)	Contains protein domain (PF01023) - S-100/CaBP type calcium binding domain	struct	1008
1877	30241217(22539, 22540)	Novel Protein sim. GBank g!3991173 sp!P31725 S109_MOUSE - CALGRANULIN B (MIGRATION INHIBITORY FACTOR-RELATED PROTEIN 14) (MRP-14) (P14) (LEUKOCYTE L1 COMPLEX HEAVY CHAIN)	Contains protein domain (PF01023) - S-100/CaBP type calcium binding domain	struct	1008

1878	80244136 (13367, 13368)	Novel Protein sim. GBank gjl173143 spl P40010 YEJ6 - YEAST - HYPOTHETICAL GTP-BINDING PROTEIN IN PML40-PAC2 INTERGENIC REGION	Contains protein domain (PF01926) - GTPase of unknown function	struct	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1017, 1022, 1023, 1024, 1026, 1029, 1031, 1034, 1037, 1039, 1044
1879	80258175 (10417, 10418)	Novel Protein sim. GBank gjl116833 spl P46681 AIP2 - YEAST - ACTIN INTERACTING PROTEIN 2		struct	1004, 1024, 1034
1880	27845416 (18229, 18230)	Novel Protein sim. GBank gjl117065 spl P33174 KIF4 - MOUSE - KINESIN-LIKE PROTEIN KIF4		struct	1010
1881	78919662 (18619, 18620)	Novel Protein sim. GBank gjl13529 spl P26039 TALIN - MOUSE - TALIN		struct	1044
1882	80204932 (22419, 22420)	Novel Protein sim. GBank gjl1362802 spl S56015 - gastric mucin - human (fragment)		struct	1004, 1006, 1022, 1024, 1034, 1039, 1054
1883	80481860 (5603, 5604)	Novel Protein sim. GBank gjl150402 spl BAA13212 - (D86976) similar to C.elegans protein (Z37093) [Homo sapiens]		struct	1006, 1010, 1012, 1024, 1027, 1031, 1038
1884	37799108 (3037, 3038)	Novel Protein sim. GBank gjl152427 spl CAB02066 - (Z779702) hypothetical protein Rv2324 [Mycobacterium tuberculosis]		struct	1012
1885	29459948 (15809, 15810)	Novel Protein sim. GBank gjl1575333 (U60416) - myr 6 myosin heavy chain [Rattus norvegicus]		struct	1038
1886	28392438 (15061, 15062)	Novel Protein sim. GBank gjl1638548 (U58971) - calmodulin-binding protein [Nicotiana tabacum]		struct	1017
1887	79112716 (14815, 14816)	Novel Protein sim. GBank gjl1928995 (U92949) - kinesin motor protein [Mus musculus]		struct	1023

1888	10870979 (16103, 16104)	Novel Protein sim. GBank gil2129582 pir S71227 - extensin 1 - Arabidopsis thaliana	struct	1010
1889	79817772 (15659, 15660)	Novel Protein sim. GBank gil2132247 pir S61015 - hypothetical protein YPL249c - yeast (Saccharomyces cerevisiae)	struct	1006, 1022, 1024, 1026, 1027
1890	8484241 (13281, 13282)	Novel Protein sim. GBank gil2224925 (AF002678) - kinesin-like protein [Arabidopsis thaliana]	struct	1022
1891	94992176 (15793, 15794)	Novel Protein sim. GBank gil2492678 sp Q12396 ARP8_YEAST - ACTIN-LIKE PROTEIN ARP8	struct	1004, 1010
1892	80450846 (1321, 1322)	Novel Protein sim. GBank gil2493435 sp Q62261 SPCO_MOUSE - SPECTRIN BETA CHAIN, BRAIN (SPECTRIN, NON-ERYTHROID BETA CHAIN) (FODRIN BETA CHAIN) (SPTBN1)	struct	1038
1893	78767589 (7949, 7950)	Novel Protein sim. GBank gil2497124 sp Q05050 YMS1_YEAST - HYPOTHETICAL 93.3 KD PROTEIN IN TAP42-ARP9 INTERGENIC REGION	struct	1008
1894	10172511 (18155, 18156)	Novel Protein sim. GBank gil2506246 sp P15508 SPCB_MOUSE - SPECTRIN BETA CHAIN, ERYTHROCYTE	struct	1038
1895	80042498 (1967, 1968)	Novel Protein sim. GBank gil266334 sp Q01042 IE68_HSVSA - IMMEDIATE-EARLY PROTEIN	struct	1053
1896	30178381 (6201, 6202)	Novel Protein sim. GBank gil2760296 dbj BAA241611 - (D87323) antigen 85C and MPT45 [Mycobacterium avium]	struct	1008
1897	95288674 (721, 722)	Novel Protein sim. GBank gil2765411 emb CAA74749 - (Y14391) GTP-binding protein [Homo sapiens]	struct	1006, 1022, 1031, 1034
1898	10236155 (21119, 21120)	Novel Protein sim. GBank gil2950347 emb CAA04492 - (AJ001038) M-protein [Mus musculus]	struct	1001

1899	88086580 (11761, 11762)	Novel Protein sim. GBank gi 3063517 (AC004594) - Ca2+ dependent activator protein for secretion; similar to D86214 (NID:gi 1398903) [Homo sapiens]	struct	1044
1900	16760037 (2763, 2764)	Novel Protein sim. GBank gi 3449356 (U92883) - slow skeletal muscle troponin T, high Mr isoform 2 [Mus musculus]	struct	1044
1901	79837459 (35665, 35666)	Novel Protein sim. GBank gi 3877745 emb CAB00100.1 - (Z75953) similar to adducin; cDNA EST EMBL:D666652 comes from this gene; cDNA EST EMBL:D70511 comes from this gene; cDNA EST yk299h7.3 comes from this gene; cDNA EST yk299n7.5 comes from this gene; cDNA EST yk315b6.3 comes fro...	struct	1001, 1008
1902	94992299 (19613, 19614)	Novel Protein sim. GBank gi 3878400 emb CAA95828 - (Z71264) predicted using Genefinder; Weak similarity to Mouse T-complex-associated-testes-expressed-1-protein (PIR Acc. No. A45841); cDNA EST EMBL:D32742 comes from this gene; cDNA EST EMBL:D33617 comes from this gene; cDNA EST...	struct	1001, 1024, 1027
1903	79208528 (1395, 1396)	Novel Protein sim. GBank gi 3914992 spl Q26264 SM41_HEMPU - 41 KD SPICULE MATRIX PROTEIN PRECURSOR (HSM41) (HPSMC)	struct	1001
1904	80434272 (20357, 20358)	Novel Protein sim. GBank gi 3929110 (AF067180) - kinesin heavy chain [Mus musculus]	struct	1038
1905	80058786 (20155, 20156)	Novel Protein sim. GBank gi 393194 (L02375) - S-antigen [Plasmoidium falciparum]	struct	1006, 1016, 1022, 1054
1906	87465840 (6363, 6364)	Novel Protein sim. GBank gi 4006920 emb CAB16815.1 - (Z99708) actin interacting protein [Arabidopsis thaliana]	struct	1019, 1022, 1025, 1038, 1058

1907	78469114 (1779, 1780)	Novel Protein sim. GBank gi 4090864 (AF023181) - low temperature requirement B protein [Listeria monocytogenes]		struct	1026
1908	79258931 (14805, 14806)	Novel Protein sim. GBank gi 4176889 sp P32843 RN12_YEAST - RNA12 PROTEIN		struct	1003, 1008, 1009
1909	17900558 (7247, 7248)	Novel Protein sim. GBank gi 4510348 gb AAD21437.1 - (AC006921) unknown protein [Arabidopsis thaliana]		struct	1044
1910	80231922 (13145, 13146)	Novel Protein sim. GBank gi 4758748 ref NP_004524.1 pMYBP - myosin-binding protein C, fast-type		struct	1004, 1024
1911	38067317 (14961, 14962)	Novel Protein sim. GBank gi 5454130 ref NP_006280.1 pTLDI - talin		struct	1022
1912	78678137 (7563, 7564)	Novel Protein sim. GBank gi 730978 sp P40414 TPM2_YEAST - TROPOMYOSIN 2		struct	1022
1913	79567530 (13017, 13018)	Novel Protein sim. GBank gi 731172 sp P17437 XP2_XENLA - SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN)		struct	1008
1914	79827013 (8545, 8546)	Novel Protein sim. GBank gi 854065 emb CAA583371 - (X83413) U88 [Human herpesvirus 6]		struct	1003
1915	20467520 (18805, 18806)			struct	1010
1916	10371483 (20687, 20688)			struct	1008
1917	78680470 (329, 330)	Novel Protein sim. GBank gi 1711602 sp P49880 SUO6_RAT - ESTROGEN SULFOTRANSFERASE, ISOFORM 6 (EST-6) (SULFOTRANSFERASE, ESTROGEN-PREFERRING) (ESTRONE SULFOTRANSFERASE)		sulfotransferase	1009, 1026, 1058

1918	78975035 (4521, 4522)	Novel Protein sim. GBank gil4758038 ref NP_004852.1 pCST - cerebroside sulfotransferase		sulfotransferase	1019, 1044
1919	25316236 (6303, 6304)	Novel Protein sim. GBank gil173146 (M27174) - carbamyl phosphate synthetase [Sacccharomyces cerevisiae]	Contains protein domain (PF00117) - Glutamine amidotransferase class-I	synthase	1009
1920	80234097 (3919, 3920)	Novel Protein sim. GBank gil1332386 (M27315) - Rattus norvegicus ATPase subunit 6 [Rattus norvegicus]	Contains protein domain (PF00119) - ATP synthase A chain	synthase	1009, 1022
1921	25256624 (1415, 1416)	Novel Protein sim. GBank gil585202 sp P3856 IGLN2_MAIZE - GLUTAMINE SYNTHETASE ROOT ISOZYME 2 (GLUTAMATE-AMMONIA LIGASE)	Contains protein domain (PF00120) - Glutamine synthetase	synthase	1003
1922	29457246 (19697, 19698)	Novel Protein sim. GBank gil140638 sp P2384 XAPR_ECOLI - XANTHOSONINE OPERON REGULATORY PROTEIN	Contains protein domain (PF00126) - Bacterial regulatory helix-turn-helix protein, lysR family	synthase	1026
1923	95287842 (265, 266)	Novel Protein sim. GBank gil3046322 (AF010139) - O-acetylserine synthase; CysE2 [Azotobacter vinelandii]	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	synthase	1006, 1008, 1010, 1012, 1016, 1022, 1024, 1034
1924	86473890 (22741, 22742)	Novel Protein sim. GBank gil114653 sp P00843 ATPH_SP10L_ATP SYNTHASE C CHAIN (LIPID-BINDING PROTEIN) (SUBUNIT III)	Contains protein domain (PF00137) - ATP synthase subunit C	synthase	1054
1925	80079063 (19869, 19870)	Novel Protein sim. GBank gil249939 sp P77883 PYRB_LACPL - ASPARTATE CARBAMOYLTRANSFERASE (ASPARTATE TRANSCARBAMYLASE) (ATCASE)	Contains protein domain (PF00185) - Aspartate/ornithine carbamoyltransferase	synthase	1034
1926	82062075 (10555, 10556)	Novel Protein sim. GBank gil117698 sp P42415 OLD_BACSU - IOLD PROTEIN	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	1006, 1016, 1022

1927	78467242 (11139, 11140)	Novel Protein sim. GBank gil1176988 sp P42415 IOLD_BACSU - IOLD PROTEIN	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	1026
1928	13084904 (15369, 15370)	Novel Protein sim. GBank gil1881244 db BA19271 - (AB001488) SIMILAR TO PYRUVATE OXIDASE AND ACETOLACTATE SYNTHASE. [Bacillus subtilis]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	1024
1929	55281880 (9735, 9736)	Novel Protein sim. GBank gil4467263 emb CAB37588 - (AL035569) acetolactate synthase [Streptomyces coelicolor]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	1029, 1040, 1053
1930	13518136 (8173, 8174)	Novel Protein sim. GBank gil3169557 gb AAC63502.1 - (AF054585) tryptophan synthase beta subunit [Streptomyces coelicolor]	Contains protein domain (PF00218) - Indole-3-glycerol phosphate synthases	synthase	1024
1931	66142987 (7849, 7850)	Novel Protein sim. GBank gil168574 sp P42464 ATPB_CORGL - ATP SYNTHASE BETA CHAIN	Contains protein domain (PF00231) - ATP synthase	synthase	1010
1932	25244094 (10711, 10712)	Novel Protein sim. GBank gil1703747 sp Q10597 ATPG_MYCTU - ATP SYNTHASE GAMMA CHAIN	Contains protein domain (PF00231) - ATP synthase	synthase	1008
1933	85518616 (11601, 11602)	Novel Protein sim. GBank gil1703747 sp Q10597 ATPG_MYCTU - ATP SYNTHASE GAMMA CHAIN	Contains protein domain (PF00231) - ATP synthase	synthase	1049
1934	78335762 (18419, 18420)	Novel Protein sim. GBank gil1703749 sp P50007 ATPG_STRLI - ATP SYNTHASE GAMMA CHAIN	Contains protein domain (PF00231) - ATP synthase	synthase	1026
1935	94134802 (8541, 8542)	Novel Protein sim. GBank gil543867 sp P26360 ATP3_IPOBA - ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL PRECURSOR	Contains protein domain (PF00231) - ATP synthase	synthase	1010
1936	795388554 (16799, 16800)	Novel Protein sim. GBank gil1168958 sp P43635 CISX_YEAST - CITRATE SYNTHASE 3	Contains protein domain (PF00285) - Citrate synthase	synthase	1022

1937	24132883 (18549, 18550)	Novel Protein sim. GBank gil53601631gb AAD2885_1 AF15701 - (AF157018) PppC [Legionella pneumophila]	Contains protein domain (PF00285) - Citrate synthase	synthase	1003
1938	80503437 (20459, 20460)	Novel Protein sim. GBank gil10760131pir A49930 - carB protein homolog - Mycobacterium bovis (strain BCG) (fragment)	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1012
1939	71761587 (8973, 8974)	Novel Protein sim. GBank gil1156261spl P25994 CARB_BACSU - CARBA MOYL-PHOSPHATE SYNTHASE, PYRIMIDINE-SPECIFIC, LARGE CHAIN (CARBA MOYL-PHOSPHATE SYNTHETASE AMMONIA CHAIN)	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1008, 1026, 1054
1940	27280026 (647, 648)	Novel Protein sim. GBank gil173146 (M27174) - carbamyl phosphate synthetase [Saccharomyces cerevisiae]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1026
1941	71199349 (20631, 20632)	Novel Protein sim. GBank gil1750387 (U81259) - carbamoylphosphate synthetase large subunit [Pseudomonas aeruginosa]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1026
1942	80071866 (7433, 7434)	Novel Protein sim. GBank gil26611433 emb CAA05020 - (AJ001805) carbamoylphosphate synthetase large subunit [Bacillus stearothermophilus]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1034, 1053
1943	79778552 (22973, 22974)	Novel Protein sim. GBank gil3873545 emb CA422122 - (AL033534) carbamoyl-phosphate synthase [Schizosaccharomyces pombe]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPsase)	synthase	1017, 1027
1944	30500812 (4667, 4668)	Novel Protein sim. GBank gil136373spl P00931 TRP_YEAST - TRYPTOPHAN SYNTHASE	Contains protein domain (PF00290) - Tryptophan synthase alpha chain	synthase	1026
1945	10312207 (11227, 11228)	Novel Protein sim. GBank gil1366604 spl P11044 TYSB_BACSU - THYMIDYLATE SYNTHASE B (TS B) (TSASE B)	Contains protein domain (PF00303) - Thymidylate synthase	synthase	1001

1946	13084260 (825, 826)	Novel Protein sim. GBank gil136608 sp P00470 TYSY_ECOLI - THYMIDYLATE SYNTHASE (TS)	Contains protein domain (PF00303) - Thymidylate synthase	synthase	1022
1947	80219835 (16559, 16560)	Novel Protein sim. GBank gil136618 sp P06785 TYSY_YEAST - THYMIDYLATE SYNTHASE (TS)	Contains protein domain (PF00303) - Thymidylate synthase	synthase	1001, 1003, 1017, 1022, 1044
1948	78762355 (16161, 16162)	Novel Protein sim. GBank gil1168574 sp P42464 ATPB_CORGL - ATP SYNTHASE BETA CHAIN	Contains protein domain (PF00306) - ATP synthase ab C terminal	synthase	1017
1949	57015995 (10415, 10416)	Novel Protein sim. GBank gil1352830 sp P49087 VATA_MAIZE - VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (V-ATPASE 69 KD SUBUNIT)	Contains protein domain (PF00306) - ATP synthase ab C terminal	synthase	1003, 1009, 1026
1950	80497309 (3965, 3966)	Novel Protein sim. GBank gil3449252 emb CA20396 - (AL031317) glucosamine--fructose-6-phosphate aminotransferase [Streptomyces coelicolor]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	synthase	1012
1951	20466338 (5285, 5286)	Novel Protein sim. GBank gil3449252 emb CA20396 - (AL031317) glucosamine--fructose-6-phosphate aminotransferase [Streptomyces coelicolor]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	synthase	1006, 1010
1952	10869112 (6803, 6804)	Novel Protein sim. GBank gil3449252 emb CA20396 - (AL031317) glucosamine--fructose-6-phosphate aminotransferase [Streptomyces coelicolor]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	synthase	1006
1953	27968471 (12353, 12354)	Novel Protein sim. GBank gil3449252 emb CA20396 - (AL031317) glucosamine--fructose-6-phosphate aminotransferase [Streptomyces coelicolor]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	synthase	1006
1954	79167910 (18871, 18872)	Novel Protein sim. GBank gil2493046 sp Q40089 ATP4_IPOBA - ATP SYNTHASE DELTA' CHAIN MITOCHONDRIAL PRECURSOR	Contains protein domain (PF00401) - ATP synthase, Delta/Epsilon chain	synthase	1031, 1050
1955	56149732 (8109, 8110)	Novel Protein sim. GBank gil145840 (M36700) - iron chelator protein [Escherichia coli]	Contains protein domain (PF00425) - chorismate binding enzyme	synthase	1003

1956	79830340 (16259, 16250)	Novel Protein sim. GBank g 2861178 dbj BAA02200 - (D12706) anthranilate synthetase [Pseudomonas aeruginosa]	Contains protein domain (PF00425) - chorismate binding enzyme	synthase	1039
1957	80092840 (2355, 2356)	Novel Protein sim. GBank g 3927835 (AC005727) - similar to Streptomyces PapA [Arabidopsis thaliana]	Contains protein domain (PF00425) - chorismate binding enzyme	synthase	1004, 1029
1958	19521859 (11055, 11056)	Novel Protein sim. GBank g 464939 sp Q06128 TRPE_SULSO_ANTHRANILATE_SYNTHASE_COMPONENT_1	Contains protein domain (PF00425) - chorismate binding enzyme	synthase	1034
1959	80414870 (13217, 13218)	Novel Protein sim. GBank g 3511285 (AF081934) - cellulose synthase [Populus alba x Populus tremula]	Contains protein domain (PF00446) - Gonadotropin-releasing hormones	synthase	1025, 1039
1960	65640665 (16605, 16606)	Novel Protein sim. GBank g 1708173 sp P54919 IHEM2_STRCO_DELTA_AMINOLEVULINIC_ACID_DEHYDRATASE (PORPHOBILINOGEN SYNTHASE) (ALAD) (ALADH)	Contains protein domain (PF00490) - Delta-aminolevulinic acid dehydratase	synthase	1016
1961	11617364 (11499, 11500)	Novel Protein sim. GBank g 171684 (J03493) - delta-aminolevulinate dehydratase [Saccharomyces cerevisiae]	Contains protein domain (PF00490) - Delta-aminolevulinic acid dehydratase	synthase	1004
1962	21659603 (3307, 3308)	Novel Protein sim. GBank g 2495166 sp P77323 IHEM2_PROFR_DELTA_AMINOLEVULINIC_ACID_DEHYDRATASE (PORPHOBILINOGEN SYNTHASE) (ALAD) (ALADH)	Contains protein domain (PF00490) - Delta-aminolevulinic acid dehydratase	synthase	1034
1963	65685418 (10477, 10478)	Novel Protein sim. GBank g 135060 sp P04712 SUS1_MAIZE - SUCROSE SYNTHASE 1 (SUCROSE-UDP GLUCOSYLTRANSFERASE 1) (SHRUNKEN-1)	Contains protein domain (PF00534) - Glycosyl transferases group 1	synthase	1049
1964	20711313 (12411, 12412)	Novel Protein sim. GBank g 2145938 pir S73015 - polyketide synthase pksF - Mycobacterium leprae	Contains protein domain (PF00550) - Phosphopantetheine attachment site	synthase	1022

1965	79840413 (10703, 10704)	Novel Protein sim. GBank gil3510629 (AF047828) - syringomycin synthetase [Pseudomonas syringae pv. syringae]	Contains protein domain (PF00550) - Phosphopantetheine attachment site	synthase	1008
1966	19881247 (21609, 21610)	Novel Protein sim. GBank gil1524206[emb CAB01994] - (Z79700) purN [Mycobacterium tuberculosis]	Contains protein domain (PF00551) - Formyl transferase	synthase	1034
1967	20262328 (8841, 8842)	Novel Protein sim. GBank gil1562542 (U68765) - 5'-phosphoribosylglycinamide transformylase [Salmonella typhimurium]	Contains protein domain (PF00551) - Formyl transferase	synthase	1022
1968	79633450 (1783, 1784)	Novel Protein sim. GBank gil136315[sp P07285 TRPD_YEAST - ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE]	Contains protein domain (PF00591) - glycosyl transferase family	synthase	1038
1969	78453670 (193, 194)	Novel Protein sim. GBank gil2952028 (U88888) - cardiolipin synthase [Bacillus firmus]	Contains protein domain (PF00614) - Phospholipase D. Active site motif	synthase	1008
1970	11084727 (10663, 10664)	Novel Protein sim. GBank gil2791611[emb CAA16099_11 - (AL021287) ligA] [Mycobacterium tuberculosis]	Contains protein domain (PF00633) - Helix-hairpin-helix motif	synthase	1004
1971	55286585 (20939, 20940)	Novel Protein sim. GBank gil1723011[sp Q10786 Y04P_MYCTU - HYPOTHETICAL 33.0 KD PROTEIN CY48.25C	Contains protein domain (PF00849) - RNA pseudouridylate synthase	synthase	1029, 1053
1972	38070114 (3443, 3444)	Novel Protein sim. GBank gil1723614[sp P50513 RCLUD_ZYMMO - RIBOSOMAL LARGE SUBUNIT PSEUDOOURIDYLATE SYNTHASE D (PSEUDOOURIDYLATE SYNTHASE) (URACIL HYDROLYASE)]	Contains protein domain (PF00849) - RNA pseudouridylate synthase	synthase	1016
1973	20617367 (12007, 12008)	Novel Protein sim. GBank gil2829541[sp P71685 RISB_MYCTU - 6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE (DMRL SYNTHASE) (LUMAZINE SYNTHASE) (RIBOFLAVIN SYNTHASE BETA CHAIN)]	Contains protein domain (PF00885) - 6,7-dimethyl-8-ribityllumazine synthase	synthase	1004, 1024
1974	79627142 (19307, 19308)	Novel Protein sim. GBank gil2076691[emb CAB08373] - (Z95150) moeB [Mycobacterium tuberculosis]	Contains protein domain (PF00899) - Thf family	synthase	1038

1975	28392902 (10833, 10834)	Novel Protein sim. GBank gil115621 spl P00907 CARA_ECOLI - CARBAMOYL-PHOSPHATE SYNTHASE SMALL CHAIN (CARBAMOYL-PHOSPHATE SYNTHETASE GLUTAMINE CHAIN)	Contains protein domain (PF00988) - Carbamoyl-phosphate synthase small chain, CPSase domain	synthase	1008
1976	65472418 (14457, 14458)	Novel Protein sim. GBank gil284621 pir B42254 - H+-transporting ATPase (EC 3.6.1.35) chain B isoform 2 - bovine (fragment)	Contains protein domain (PF01038) - ATP synthase beta chain, carboxyl terminus	synthase	1004, 1023
1977	79183610 (1145, 1146)	Novel Protein sim. GBank gil548438 spl P06978 PGSA_ECOLI - CDP-DIACYLGLYCEROL-GLYCEROL-3-PHOSPHATE 3-PHOSPHATIDYLTRANSFERASE (PHOSPHATIDYL GLYCEROPHOSPHATE SYNTHASE) (PGP SYNTHASE)	Contains protein domain (PF01066) - CDP-alcohol phosphatidyltransferase	synthase	1027
1978	35605314 (17507, 17508)	Novel Protein sim. GBank gil4102219 (AF011408) - type II 3-dehydroquinase [Aeromonas salmonicida subsp. salmonicida]	Contains protein domain (PF01220) - Dehydroquinase class II	synthase	1053
1979	78513848 (4251, 4252)	Novel Protein sim. GBank gil3122401 spl P06219 MURE_MYCTU - UDP-N-ACETYL MURAMOYL ALANYL-D-GLUTAMATE-2,6-DIAMINOPIMELATE LIGASE (UDP-N-ACETYL MURAMYL-TRIPEPTIDE SYNTHETASE)	Contains protein domain (PF01225) - Mur ligase family	synthase	1000, 1008, 1026, 1054
1980	80502410 (1751, 1752)	Novel Protein sim. GBank gil3122401 spl P07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)	Contains protein domain (PF01411) - tRNA synthetases class II (A)	synthase	1003, 1006, 1008, 1010, 1012, 1034
1981	71762749 (22555, 22556)	Novel Protein sim. GBank gil4539134 emb CA39580.1 - (AL049478) putative pseudouridine synthase [Mycobacterium leprae]	Contains protein domain (PF01479) - S4 domain	synthase	1008
1982	79588046 (14491, 14492)	Novel Protein sim. GBank gil231772 spl P30538 CHS1_USTMA - CHITIN SYNTETASE 1 (CHITIN-UDP ACETYL-GLUCOSAMINYL TRANSFERASE 1)	Contains protein domain (PF01644) - Chitin synthase	synthase	1034

1983	79608332 (18113, 18114)	Novel Protein sim. GBank gil119832 sp P19097 FAS2_YEAST - FATTY ACID SYNTHASE, SUBUNIT ALPHA [CONTAINS: EC 1.1.1.100; EC 2.3.1.41]	Contains protein domain (PF01646) - Holo-acyl carrier protein synthase	synthase	1006, 1038
1984	95295611 (2345, 2346)	Novel Protein sim. GBank gil41839 sp P32610 VATD_YEAST - VACUOLAR ATP SYNTHASE SUBUNIT D (V-ATPASE D SUBUNIT)	Contains protein domain (PF01813) - ATP synthase subunit D	synthase	1001, 1025
1985	86473038 (6493, 6494)	Novel Protein sim. GBank gil1001348 dbj BAA10835 - (D64006) hypothetical protein [Synechocystis sp.]		synthase	1029
1986	17673579 (12191, 12192)	Novel Protein sim. GBank gil100529 pir S19261 - H+-transporting ATP synthase (EC 3.6.1.34) alpha chain - common sunflower mitochondrion		synthase	1008
1987	18597085 (18765, 18766)	Novel Protein sim. GBank gil1073121 pir S55589 - D-nopaline dehydrogenase (EC 1.5.1.19) OoxA - Agrobacterium tumefaciens		synthase	1039
1988	13519586 (4257, 4258)	Novel Protein sim. GBank gil1074100 pir H64090 - formylglycineamide ribonucleotide synthetase (purL) homolog - Haemophilus influenzae (strain Rd KW20)		synthase	1024
1989	30662558 (3549, 3550)	Novel Protein sim. GBank gil1074945 pir E64106 - isopropylmalate synthase (leuA) homolog - Haemophilus influenzae (strain Rd KW20)		synthase	1026
1990	21630256 (5979, 5980)	Novel Protein sim. GBank gil1075536 pir S52226 - cobQ protein - Rhodobacter capsulatus		synthase	1038
1991	10852831 (11861, 11862)	Novel Protein sim. GBank gil1075778 pir JX0340 - gramicidin S synthase 2 - Bacillus brevis		synthase	1044
1992	23330118 (11595, 11596)	Novel Protein sim. GBank gil1075998 pir S555505 - fatty-acid synthase (EC 2.3.1.85) - Brevibacterium ammoniagenes		synthase	1026

1993	79186424 (17743, 17744)	Novel Protein sim. GBank gi 114135 sp P08205 ARGA_ECOLI - AMINO- ACID ACETYLTRANSFERASE (N- ACETYLGLUTAMATE SYNTHASE) (AGS)		synthase	1027
1994	27366289 (8231, 8232)	Novel Protein sim. GBank gi 114166 sp P08566 ARO1_YEAST - PENTAFUNCTIONAL AROM POLYPEPTIDE [CONTAINS: 3-DEHYDROQUINATE SYNTHASE ; 3-DEHYDROQUINATE DEHYDRATASE (3- DEHYDROQUINASE); SHIKIMATE 5- DEHYDROGENASE; SHIKIMATE KINASE; EPSP SYNTHASE]		synthase	1017
1995	78835590 (18685, 18686)	Novel Protein sim. GBank gi 114186 sp P28777 AROC_YEAST - CHORISMATE SYNTHASE (5- ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE)		synthase	1008, 1009
1996	80204223 (14795, 14796)	Novel Protein sim. GBank gi 114263 sp P22106 ASN_B_ECOLI - ASPARAGINE SYNTHETASE B (GLUTAMINE- HYDROLYZING)		synthase	1004, 1017, 1023, 1044
1997	79262179 (7673, 7674)	Novel Protein sim. GBank gi 114291 sp P09034 ASSY_RAT - ARGININOSUCCINATE SYNTHASE (CITRULLINE--ASPARTATE LIGASE)		synthase	1003, 1008, 1023
1998	78400485 (14995, 14996)	Novel Protein sim. GBank gi 1144522 U34957 - phosphoribosylaminimidazolesuccinocarboxamid e synthase [Mycobacterium tuberculosis]		synthase	1038
1999	11090659 (20717, 20718)	Novel Protein sim. GBank gi 1144522 U34957 - phosphoribosylaminimidazolesuccinocarboxamid e synthase [Mycobacterium tuberculosis]		synthase	1006
2000	55281689 (1253, 1254)	Novel Protein sim. GBank gi 115001 sp P19206 BACSH_BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	1029, 1053

2001	80025977 (10059, 10060)	Novel Protein sim. GBank gll1150011sp P19206 BIOB_BACSH - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	1004, 1006, 1010, 1024, 1025, 1034
2002	85801295 (14937, 14938)	Novel Protein sim. GBank gll1150121sp P22822 BLOW_BACSH - 6- CARBOXYHEXANOATE-COA LIGASE (PIMELOYL-COA SYNTHASE)		synthase	1012, 1021, 1025
2003	78794886 (21047, 21048)	Novel Protein sim. GBank gll1152081sp P07245 C1TC_YEAST - C-1- TETRAHYDROFOLATE SYNTHASE, CYTOPLASMIC (C1-THF SYNTHASE) (METHYLENETETRAHYDROFOLATE DEHYDROGENASE / METHENYL-TETRAHYDROFOLATE CYCLOHYDROLASE / FORMYL-TETRAHYDROFOLATE SYNTHETASE		synthase	1044
2004	79875269 (4867, 4868)	Novel Protein sim. GBank gll1163821sp P14180 CHS2_YEAST - CHITIN SYNTHASE 2 (CHITIN-UDP ACETYL- GLUCOSAMINYL TRANSFERASE 2)		synthase	1001
2005	80429039 (21169, 21170)	Novel Protein sim. GBank gll1163821sp P14180 CHS2_YEAST - CHITIN SYNTHASE 2 (CHITIN-UDP ACETYL- GLUCOSAMINYL TRANSFERASE 2)		synthase	1004, 1006, 1008, 1010, 1012, 1022, 1024, 1025, 1026, 1034, 1038, 1044
2006	79875665 (13089, 13090)	Novel Protein sim. GBank gll11685021sp P43905 AROA_LACLA - 3- PHOSPHOSHIKIMATE 1- CARBOXYVINYL TRANSFERASE (5- ENOLPYRUVYL SHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP SYNTHASE)		synthase	1039
2007	95267618 (6109, 6110)	Novel Protein sim. GBank gll11685741sp P42464 ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	1006, 1010, 1012, 1016, 1022, 1038

2008	80488983 (13539, 13540)	Novel Protein sim. GBank g!1168574 sp!P42464 ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	1003, 1006, 1008, 1010, 1016, 1022, 1023, 1038, 1049
2009	27976936 (10489, 10490)	Novel Protein sim. GBank g!1169385 sp!P43813 DNLJ_HAEIN - DNA LIGASE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (NAD+))		synthase	1022
2010	23284776 (3949, 3950)	Novel Protein sim. GBank g!1170637 sp!P45251 KDSA_HAEIN - 2- DEHYDRO-3-DEOXYPHOSPHOACONATE ALDOLASE (PHOSPHO-2-DEHYDRO-3- DEOXYOCTONATE ALDOLASE) (3-DEOXY-D- MANNO-OCTULOSONIC ACID 8-PHOSPHATE SYNTHETASE) (KDO 8-P SYNTHASE)		synthase	1008
2011	71462882 (11041, 11042)	Novel Protein sim. GBank g!1170739 sp!P44463 LIP_A_HAEIN - LIPOIC ACID SYNTHETASE (LIP-SYN) (LIPOATE SYNTHASE)		synthase	1008
2012	29250624 (18305, 18306)	Novel Protein sim. GBank g!1170917 sp!P44612 MEND_HAEIN - 2- SUCCINYL-6-HYDROXY-2,4- CYCLOHEXADIENE-1-CARBOXYLATE SYNTHASE (SHHC) (SYNTHASE) / 2- OXOGLUTARATE DECARBOXYLASE (ALPHA- KETOGLUTARATE DECARBOXYLASE) (KDC)		synthase	1008, 1009, 1031
2013	13090009 (22423, 22424)	Novel Protein sim. GBank g!1171129 (U24657) - saframycin Mx1 synthetase A [Myxococcus xanthus]		synthase	1027
2014	78779851 (3489, 3490)	Novel Protein sim. GBank g!1173422 sp!P4390 ISELA_HAEIN - L-SERYL- TRNA(SER) SELENIUM TRANSFERASE (CYSTEINYLN-TRNA(SER) SELENIUM TRANSFERASE) (SELENOCYSTEINE SYNTHASE) (SELENOCYSTEINYLN-TRNA(SER) SYNTHASE)		synthase	1022, 1026

2015	80221490 (3003, 3004)	Novel Protein sim. GBank gil1174971 sp P41807 VM13_YEAST - VACUOLAR ATP SYNTHASE 54 KD SUBUNIT (VATPASE 54 KD SUBUNIT)		synthase	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1017, 1022, 1023, 1024, 1026, 1027, 1031, 1034, 1039, 1044
2016	80491798 (3421, 3422)	Novel Protein sim. GBank gil1174971 sp P41807 VM13_YEAST - VACUOLAR ATP SYNTHASE 54 KD SUBUNIT (VATPASE 54 KD SUBUNIT)		synthase	1001, 1004, 1006, 1009, 1012, 1017, 1022, 1023, 1024, 1025, 1027, 1031, 1034, 1038
2017	80077581 (14101, 14102)	Novel Protein sim. GBank gil118234 sp P21829 MASZ_YEAST - MALATE SYNTHASE 2, GLYOXYOSOMAL		synthase	1003, 1022, 1023, 1026, 1034
2018	28980368 (2359, 2360)	Novel Protein sim. GBank gil118823 sp P14020 DPM1_YEAST - DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (DOLICHOL-PHOSPHATE MANNOSE SYNTHASE) (DOLICHYL-PHOSPHATE BETA-D-MANNOSYLTRANSFERASE)		synthase	1004, 1044
2019	78903290 (16881, 16882)	Novel Protein sim. GBank gil119837 sp P19096 FAS_MOUSE - FATTY ACID SYNTHASE [CONTAINS: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]		synthase	1039
2020	71811275 (13239, 13240)	Novel Protein sim. GBank gil1227027 sp P06106 MT17_YEAST - O-ACETYLHOMOSERINE SULPHYDRYLASE (OAH SULPHYDRYLASE) / O-ACETYLSERINE SULPHYDRYLASE (OAS SULPHYDRYLASE)		synthase	1008

2021	94140232 (15483, 154B4)	Novel Protein sim. GBank gii128480 sp P08633 NODM_RHILV - GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING) (GFAT) (NODULATION PROTEIN M)		synthase	1010
2022	27974676 (5777, 5778)	Novel Protein sim. GBank gii131619 sp P07244 PUR2_YEAST - PHOSPHORIBOSYLAMINE--GLYCINE LIGASE (GARS) (GLYCINAMIDE RIBONUCLEOTIDE SYNTHETASE) / (PHOSPHORIBOSYL GLYCINAMIDE SYNTHETASE) / PHOSPHORIBOSYLFORMYL GLYCINAMIDINE CYCLO-LIGASE (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) ...		synthase	1022
2023	19881189 (7303, 7304)	Novel Protein sim. GBank gii131619 sp P07244 PUR2_YEAST - PHOSPHORIBOSYLAMINE--GLYCINE LIGASE (GARS) (GLYCINAMIDE RIBONUCLEOTIDE SYNTHETASE) / (PHOSPHORIBOSYL GLYCINAMIDE SYNTHETASE) / PHOSPHORIBOSYLFORMYL GLYCINAMIDINE CYCLO-LIGASE (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) ...		synthase	1034
2024	27978111 (2269, 2270)	Novel Protein sim. GBank gii131647 sp P15254 PUR1_ECOLI - PHOSPHORIBOSYLFORMYL GLYCINAMIDINE SYNTHASE (FGAM SYNTHASE) (FORMYL GLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT)		synthase	1006

2025	88465195 (3967, 3968)	Novel Protein sim. GBank gi 131647 sp P15254 PURL_ECOLI-PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE (FGAM SYNTHASE) (FORMYLGLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT)		synthase	1029
2026	20438280 (14343, 14344)	Novel Protein sim. GBank gi 131647 sp P15254 PURL_ECOLI-PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE (FGAM SYNTHASE) (FORMYLGLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT)		synthase	1010
2027	27981637 (17995, 17996)	Novel Protein sim. GBank gi 131647 sp P15254 PURL_ECOLI-PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE (FGAM SYNTHASE) (FORMYLGLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT)		synthase	1006
2028	37031235 (1693, 1694)	Novel Protein sim. GBank gi 131732 sp P28593 PYRG_AZOB - CTP SYNTHASE (UTP-AMMONIA LIGASE) (CTP SYNTHETASE)		synthase	1038
2029	49318395 (1035, 1036)	Novel Protein sim. GBank gi 133985 dbj BAA12741 - (D85230) large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]		synthase	1036
2030	20293372 (5319, 5320)	Novel Protein sim. GBank gi 133995 dbj BAA12742 - (D85230) small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]		synthase	1034
2031	30652832 (14095, 14096)	Novel Protein sim. GBank gi 1351812 sp P48570 HOOSC_YEAST-HOMOCITRATE SYNTHASE, CYTOSOLIC ISOZYME		synthase	1008

2032	25307355 (15491, 15492)	Novel Protein sim. GBank g!1351812 sp P48570 HOSC_YEAST - HOMOCITRATE SYNTHASE, CYTOSOLIC ISOZYME		synthase	1022
2033	17900519 (291, 292)	Novel Protein sim. GBank g!135816 sp P16120 THRC_YEAST - THREONINE SYNTHASE		synthase	1039
2034	32152832 (8083, 8084)	Novel Protein sim. GBank g!136271 sp P06361 TRPB_BRELA - TRYPTOPHAN SYNTHASE BETA CHAIN		synthase	1039
2035	20436431 (5703, 5704)	Novel Protein sim. GBank g!136326 sp P20579 TRPE_PSEPU - ANTHRANILATE SYNTHASE COMPONENT I		synthase	1024
2036	80574455 (7099, 7100)	Novel Protein sim. GBank g!11363705 pir SS5314 - hypothetical protein YLR126c - yeast (Saccharomyces cerevisiae)		synthase	1058
2037	80473427 (20157, 20158)	Novel Protein sim. GBank g!146168 (J01617) - glutamyl-tRNA synthetase [Escherichia coli]		synthase	1006, 1010, 1012, 1022, 1023, 1024, 1027
2038	65459357 (5207, 5208)	Novel Protein sim. GBank g!1651762 dbj BA16690 - (D90899) geranylgeranyl pyrophosphate synthase [Synechocystis sp.]		synthase	1054
2039	80222993 (11723, 11724)	Novel Protein sim. GBank g!1702453 sp P22768 ASSY_YEAST - ARGININO-SUCCINATE SYNTHASE (CITRULLINE--ASPARTATE LIGASE)		synthase	1004, 1006
2040	20289880 (4793, 4794)	Novel Protein sim. GBank g!1703652 sp Q10592 ATPA_MYCTU - ATP SYNTHASE ALPHA CHAIN		synthase	1034
2041	71049212 (5999, 6000)	Novel Protein sim. GBank g!1705464 sp P53557 BIOB_BACSU - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	1029, 1050
2042	8239813 (1217, 1218)	Novel Protein sim. GBank g!1705465 sp P32451 BIOB_YEAST - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	1022

2043	80502400 (559, 560)	Novel Protein sim. GBank gil1708241 spl P54839 HMC5_YEAST - HYDROXYMETHYLGLUTARYL-COA SYNTHASE (HMG-COA SYNTHASE) (3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE)		synthase	1004, 1006, 1010, 1012, 1022, 1024, 1031, 1034, 1038
2044	80236746 (2089, 2090)	Novel Protein sim. GBank gil1708241 spl P54839 HMC5_YEAST - HYDROXYMETHYLGLUTARYL-COA SYNTHASE (HMG-COA SYNTHASE) (3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE)		synthase	1001, 1006, 1022
2045	28486030 (223, 224)	Novel Protein sim. GBank gil1729955 spl Q10610 THRC_MYCTU - PROBABLE THREONINE SYNTHASE		synthase	1017
2046	82090170 (12689, 12690)	Novel Protein sim. GBank gil1730280 spl P51618 GCH2_PHOPO - GTP CYCLOHYDROLASE II (RIBIV)		synthase	1016
2047	17286565 (1787, 1788)	Novel Protein sim. GBank gil173045 (K01386) - anthranilate synthase Component II:indole-3-glycerol phosphate synthase [TRP3] [Saccharomyces cerevisiae]		synthase	1017
2048	11689123 (12667, 12668)	Novel Protein sim. GBank gil173144 (J04711) - carbamylase phosphate synthetase-aspartate transcarbamylase (EC 2.1.3.2) [Saccharomyces cerevisiae]		synthase	1034
2049	20710589 (17055, 17056)	Novel Protein sim. GBank gil1750397 (U81261) - glutamate synthase large subunit [Pseudomonas aeruginosa]		synthase	1006
2050	55281693 (2705, 2706)	Novel Protein sim. GBank gil1781087 emb CAB06231 - (Z83864) gitB [Mycobacterium tuberculosis]		synthase	1029, 1053
2051	55281886 (10183, 10184)	Novel Protein sim. GBank gil1781097 emb CAB06231 - (Z83864) gitB [Mycobacterium tuberculosis]		synthase	1029, 1053

2052	20438797 (12047, 12048)	Novel Protein sim. GBank gi 17810971emb CAB06231 - (Z83864) <i>gltB</i> [Mycobacterium tuberculosis]		synthase	1024
2053	21428744 (15897, 15898)	Novel Protein sim. GBank gi 17810971emb CAB06231 - (Z83864) <i>gltB</i> [Mycobacterium tuberculosis]		synthase	1006
2054	78749414 (12137, 12138)	Novel Protein sim. GBank gi 1814403 (U84889) - methionine synthase [Mesembryanthemum crystallinum]		synthase	1008, 1009, 1022, 1026
2055	20465277 (8919, 8920)	Novel Protein sim. GBank gi 1850588 (U88301) - CTP synthetase [Mycobacterium bovis]		synthase	1010
2056	70336441 (4381, 4382)	Novel Protein sim. GBank gi 1877321emb CAB07093 - (Z92771) <i>rmlD</i> [Mycobacterium tuberculosis]		synthase	1049
2057	80105823 (19833, 19834)	Novel Protein sim. GBank gi 19063781emb CAA67249 - (X98690) <i>Pristinamycin I synthase 3 and 4</i> [Streptomyces <i>pristinaespiralis</i>]		synthase	1013
2058	77592724 (4407, 4408)	Novel Protein sim. GBank gi 19229201emb CAA70871 - (Y09700) <i>rpfB</i> [Xanthomonas campestris]		synthase	1008
2059	21658756 (5163, 5164)	Novel Protein sim. GBank gi 1929513 (U64318) - ATP synthase subunit beta [Moorella thermoacetica]		synthase	1010
2060	34394294 (22731, 22732)	Novel Protein sim. GBank gi 20766391emb CAB08373 - (Z95150) <i>moeB</i> [Mycobacterium tuberculosis]		synthase	1026
2061	85520648 (19915, 19916)	Novel Protein sim. GBank gi 2104413emb CAB08713 - (Z95390) <i>otsA</i> [Mycobacterium tuberculosis]		synthase	1009, 1022, 1049
2062	66727444 (21269, 21270)	Novel Protein sim. GBank gi 2104413emb CAB08713 - (Z95390) <i>otsA</i> [Mycobacterium tuberculosis]		synthase	1010, 1022, 1049
2063	20467298 (13383, 13384)	Novel Protein sim. GBank gi 2114017emb CAB08966 - (Z95558) <i>menD</i> [Mycobacterium tuberculosis]		synthase	1010

2064	65704247 (19201, 19202)	Novel Protein sim. GBank g 2114017 emb CAB08966 - (Z955558) menD [Mycobacterium tuberculosis]		synthase	1010
2065	79846083 (8511, 8512)	Novel Protein sim. GBank g 2125896 emb CAA73511 - (Y13070) folylpolyglutamate synthase [Streptomyces coelicolor]		synthase	1034
2066	80498648 (15931, 15932)	Novel Protein sim. GBank g 2125896 emb CAA73511 - (Y13070) folylpolyglutamate synthase [Streptomyces coelicolor]		synthase	1012
2067	16316433 (7773, 7774)	Novel Protein sim. GBank g 2129478 pir S51939 - chitinase (EC 3.2.1.14) precursor - beet		synthase	1034
2068	23308794 (163581, 16582)	Novel Protein sim. GBank g 2129675 pir S60222 - probable chlorophyll synthetase G4 - Arabidopsis thaliana		synthase	1053
2069	27362136 (11837, 11838)	Novel Protein sim. GBank g 2131263 pir S61041 - GLT1 protein - yeast (Saccharomyces cerevisiae)		synthase	1009
2070	80230544 (3989, 3990)	Novel Protein sim. GBank g 2131575 pir S64067 - hypothetical protein YGL063w - yeast (Saccharomyces cerevisiae)		synthase	1004, 1006, 1022, 1024, 1027, 1034
2071	8911723 (1883, 1884)	Novel Protein sim. GBank g 2244683 emb CAA6913 - (Y07786) ORF43x9 [Vibrio cholerae]		synthase	1044
2072	16504667 (13025, 13026)	Novel Protein sim. GBank g 2266762 emb CAA67877 - (X99521) aspartate aminotransferase [Thermus aquaticus]		synthase	1008
2073	79245632 (7643, 7644)	Novel Protein sim. GBank g 2314156 gb AAD08057.1 - (AE000609) dihydrodipicolinate synthetase (dapA) [Helicobacter pylori 26695]		synthase	1039
2074	27978192 (16971, 16972)	Novel Protein sim. GBank g 2318069 (AF012246) - carbamoyl-phosphate synthetase subunit A [Salmonella typhi]		synthase	1022

2076	27980065 (8289, 8290)	Novel Protein sim. G-Bank gi 2342614 emb CAB11385 - (Z98741) lipoc acid synthetase [Mycobacterium leprae]		synthase	1006
2076	10152983 (4875, 4876)	Novel Protein sim. G-Bank gi 2492788 sp Q12122 HOSM_YEAST - PROBABLE HOMOCITRATE SYNTHASE, MITOCHONDRIAL ISOZYME PRECURSOR		synthase	1003
2077	17940458 (20671, 20672)	Novel Protein sim. G-Bank gi 2492885 sp P77495 PRPE_ECOLI - PRPE PROTEIN		synthase	1025, 1039
2078	80078413 (20085, 20086)	Novel Protein sim. G-Bank gi 2492983 sp Q44093 AROG_AMYME - PHOSPHO-2-DIHYDRO-3-DEOXYHEPTONATE ALDOLASE (PHOSPHO-2-KETO-3-DEOXYHEPTONATE ALDOLASE) (DAHP SYNTHETASE) (3-DEOXY-D-ARABINO-HEPTULOSONATE 7-PHOSPHATE SYNTHASE)		synthase	1034
2079	78929520 (15747, 15748)	Novel Protein sim. G-Bank gi 2493129 sp Q43432 VAT1_GOSHI - VACUOLAR ATP SYNTHASE SUBUNIT B ISOFORM 1 (V-ATPASE B SUBUNIT)		synthase	1003, 1008
2080	80216096 (20475, 20476)	Novel Protein sim. G-Bank gi 2494764 sp Q50729 GUAA_MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)		synthase	1004, 1010
2081	66694923 (3165, 3166)	Novel Protein sim. G-Bank gi 2498155 sp Q59280 ARGJ_CORG1 - GLUTAMATE N-ACETYLTRANSFERASE (ORNITHINE ACETYLTRANSFERASE) (ORNITHINE TRANSACETYLASE)(OATASE)		synthase	1004, 1009, 1012, 1025, 1031
2082	27807621 (10039, 10040)	Novel Protein sim. G-Bank gi 2498415 sp Q04952 GLS3_YEAST - PUTATIVE 1,3-BETA-GLUCAN SYNTHASE COMPONENT (1,3-BETA-D-GLUCAN-UDP GLUCOSYLTRANSFERASE)		synthase	1022, 1024

2083	11762182 (691, 692)	Novel Protein sim. G Bank gi 2498853 sp P72216 LPXB_PROMI - LIPID-A-DISACCHARIDE SYNTHASE		synthase	1022
2084	10325170 (20657, 20658)	Novel Protein sim. G Bank gi 2498845 sp Q12074 SPEE_YEAST-SPERMIDINE SYNTHASE (PUTRESCINE AMINOPROPYLTRANSFERASE) (SPDSY)		synthase	1025
2085	80481640 (10791, 10792)	Novel Protein sim. G Bank gi 2500947 sp P56069 METB_HELPY-CYSTATHIONINE GAMMA-SYNTHASE (CGS) (O-SUCCINYLHOMOSERINE (THIOL)-LYASE)		synthase	1012
2086	78944492 (20771, 20772)	Novel Protein sim. G Bank gi 2501152 sp Q42558 THRC_SCHPO - THREONINE SYNTHASE		synthase	1003, 1017, 1022, 1039
2087	39505065 (20805, 20806)	Novel Protein sim. G Bank gi 2501186 sp Q59925 FTHS_STRMU - FORMATE-TETRAHYDROFOLATE LIGASE (FORMYL-TETRAHYDROFOLATE SYNTHETASE) (FHS) (FTHFS)		synthase	1004
2088	20177075 (17833, 17834)	Novel Protein sim. G Bank gi 2501680 sp P75485 Y209_MYCPN - HYPOTHETICAL PROTEIN MG209 HOMOLOG		synthase	1010
2089	80255958 (4747, 4748)	Novel Protein sim. G Bank gi 2506361 sp P04819 DNL1_YEAST - DNA LIGASE I (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (ATP))		synthase	1001, 1003, 1006, 1017, 1022
2090	17896879 (10819, 10820)	Novel Protein sim. G Bank gi 2506362 sp P15042 DNLJ_ECOLI - DNA LIGASE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE (NAD+))		synthase	1044
2091	66397533 (2091, 2092)	Novel Protein sim. G Bank gi 2507167 sp P23538 PPSA_ECOLI - PHOSPHOENOLPYRUVATE SYNTHASE (PYRUVATE, WATER DIKINASE) (PEP SYNTHASE)		synthase	1008

2092	39710251 (5867, 5868)	Novel Protein sim. GBank gil2507436[sp]P13188ISYQ_YEAST - GLUTAMINYL-TRNA SYNTHETASE (GLUTAMINE-TRNA LIGASE) (GLNRS)		synthase	1004, 1008
2093	25324015 (6195, 6196)	Novel Protein sim. GBank gil2522212 (AF023464) - fengycin synthetase FenA [Bacillus substillis]		synthase	1008
2094	80238549 (7403, 7404)	Novel Protein sim. GBank gil2582531 (AF026444) - 2-isopropylmalate synthase [Streptomyces coelicolor]		synthase	1001, 1006, 1008, 1022, 1023, 1025, 1031, 1039, 1044
2095	20727751 (9369, 9370)	Novel Protein sim. GBank gil2613108 (AF030554) - class V chitin synthase [Ustilago maydis]		synthase	1006
2096	11688569 (17925, 17926)	Novel Protein sim. GBank gil2625092 (AF030520) - argininosuccinate synthetase [Corynebacterium glutamicum]		synthase	1031
2097	11688044 (21279, 21280)	Novel Protein sim. GBank gil2625092 (AF030520) - argininosuccinate synthetase [Corynebacterium glutamicum]		synthase	1006
2098	78936953 (22387, 22388)	Novel Protein sim. GBank gil2665938 (AF036246) - putative F1-ATP synthase beta subunit [Rickettsia prowazekii]		synthase	1044
2099	37052303 (4195, 4196)	Novel Protein sim. GBank gil2791495[emb CAA16032] - (AL021246) hypothetical protein Rv2455c [Mycobacterium tuberculosis]		synthase	1050
2100	80065985 (8261, 8262)	Novel Protein sim. GBank gil2791611[emb CAA16099.1] - (AL021287) liga [Mycobacterium tuberculosis]		synthase	1022
2101	46571653 (10985, 10986)	Novel Protein sim. GBank gil2791611[emb CAA16099.1] - (AL021287) liga [Mycobacterium tuberculosis]		synthase	1029
2102	88094444 (4017, 4018)	Novel Protein sim. GBank gil2808807[emb CAA04607.1] - (AJ001206) putative trehalose synthase [Streptomyces coelicolor]		synthase	1006, 1010, 1016, 1024, 1039

2103	13527675 (18635, 18636)	Novel Protein sim. GBank gil28110331sp O05314 GLGC_MYCTU - GLUCOSE-1-PHOSPHATE ADENYLTRANSFERASE (ADP-GLUCOSE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHORYLASE)		synthase	1027
2104	79104017 (5109, 5110)	Novel Protein sim. GBank gil28333851sp Q43134 UGST_SORBI - GRANULE BOUND GLYCOGEN (STARCH) SYNTHASE PRECURSOR		synthase	1003, 1008, 1017, 1023, 1039, 1044
2105	10223322 (451, 452)	Novel Protein sim. GBank gil285222 pir B42790 - cystathionine beta-synthase (EC 4.2.1.22) type III - rat.		synthase	1006
2106	20467259 (6549, 6550)	Novel Protein sim. GBank gil2894166 emb CAA11773.1 - (AJ223998) PCZA36.1.18 [Amycolatopsis orientalis]		synthase	1010
2107	80502756 (17949, 17950)	Novel Protein sim. GBank gil2909459 emb CAA17347 - (AL021929) cobQ [Mycobacterium tuberculosis]		synthase	1006, 1012
2108	20469055 (8313, 8314)	Novel Protein sim. GBank gil2909522 emb CAA17404 - (AL021932) ufaA1 [Mycobacterium tuberculosis]		synthase	1024
2109	16332371 (20029, 20030)	Novel Protein sim. GBank gil2916935 emb CAA17672.1 - (AL022004) moae2 [Mycobacterium tuberculosis]		synthase	1004
2110	27977735 (9135, 9136)	Novel Protein sim. GBank gil2950419 emb CAA17864.1 - (AL022076) pks13 [Mycobacterium tuberculosis]		synthase	1006
2111	78468967 (10767, 10768)	Novel Protein sim. GBank gil2952529 gb AAC05774.1 - (AF051356) D-alanine-D-alanyl carrier protein ligase [Streptococcus mutans]		synthase	1026
2112	84453012 (21755, 21756)	Novel Protein sim. GBank gil2983778 (AE000736) - hypothetical protein [Aquifex aeolicus]		synthase	1049

2113 1024)	20708860 (1023, 1024)	Novel Protein sim. GBank gi 3024480 sp P71827 PUR2_MYCTU - PHOSPHORIBOSYLAMINE--GLYCINE LIGASE (GARS) (GLYCINAMIDE RIBONUCLEOTIDE SYNTETASE) (PHOSPHORIBOSYLGLYCINAMIDE SYNTETASE)		synthase	1022
2114	87569779 (3027, 3028)	Novel Protein sim. GBank gi 3024486 sp Q50144 PUR2_MYCIE - PHOSPHORIBOSYLAMINE--GLYCINE LIGASE (GARS) (GLYCINAMIDE RIBONUCLEOTIDE SYNTETASE) (PHOSPHORIBOSYLGLYCINAMIDE SYNTETASE)		synthase	1004
2115	10222707 (14561, 14562)	Novel Protein sim. GBank gi 3024981 sp P7786 YBDL_ECOLI - HYPOTHETICAL AMINOTRANSFERASE YBDL		synthase	1023
2116	17707207 (1987, 1988)	Novel Protein sim. GBank gi 3 08053 (AF056326) - myo-inositol 1-phosphate synthase; INO1 [Zea mays]		synthase	1017
2117	94240031 (15413, 15414)	Novel Protein sim. GBank gi 3 08053 (AF056326) - myo-inositol 1-phosphate synthase; INO1 [Zea mays]		synthase	1044
2118	23302166 (14693, 14694)	Novel Protein sim. GBank gi 3122391 sp O33465 METH_PSEPU - 5- METHYL TETRAHYDROFOLATE-- HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 DEPENDENT ISOZYME) (MS)		synthase	1026
2119	80448468 (20721, 20722)	Novel Protein sim. GBank gi 3123265 sp P53167 PUS2_YEAST - PSEUDOOURIDYLATE SYNTHASE 2 (PSEUDOOURIDYLATE SYNTHASE 2)		synthase	1001, 1004, 1006, 1008, 1009, 1010, 1022, 1023, 1024, 1027, 1031, 1034, 1038, 1054

2120	20707264 (665, 666)	Novel Protein sim. GBank gj 3219785 spl P87131 CYSK_SCHPO - PUTATIVE CYSTEINE SYNTHASE (O-ACETYL SERINE SULFHYDRYLASE) (O-ACETYL SERINE (THIOL)-LYASE) [CSASE]		synthase	1006
2121	19642042 (12151, 12152)	Novel Protein sim. GBank gj 3287773 spl P73538 BIOB_SYN3 - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	1010
2122	20437161 (3999, 4000)	Novel Protein sim. GBank gj 3334257 spl Q006584 MECTU_5-METHYL TETRAHYDROPTEROYLTIGLUTAMATE -HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE)		synthase	1010
2123	32168531 (2053, 2054)	Novel Protein sim. GBank gj 3334472 spl P80877 MECTU_BACSU - 5-METHYL TETRAHYDROPTEROYLTIGLUTAMATE -HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE) (SUPEROXIDE-INDUCIBLE PROTEIN 9) (SO19)		synthase	1012
2124	20456464 (16825, 18826)	Novel Protein sim. GBank gj 3410887 dbj BA32241 - (AB006850) decaprenyl diphosphate synthase [Gluconobacter suboxydans]		synthase	1010
2125	86475381 (10585, 10586)	Novel Protein sim. GBank gj 3449252 emb CAA20396 - (AL031317) glucosamine-fructose-6-phosphate aminotransferase [Streptomyces coelicolor]		synthase	1025, 1048
2126	11112444 (7713, 7714)	Novel Protein sim. GBank gj 3510629 (AF047828) - syringomycin synthetase [Pseudomonas syringae pv. syringae]		synthase	1006

2127	88094584 (7729, 7730)	Novel Protein sim. GBank gj 3510629 (AF047828) - syringomycin synthetase [Pseudomonas syringae pv. syringae]		synthetase	1004
2128	10064432 (8669, 8670)	Novel Protein sim. GBank gj 3510629 (AF047828) - syringomycin synthetase [Pseudomonas syringae pv. syringae]		synthetase	1017
2129	39559724 (11179, 11180)	Novel Protein sim. GBank gj 3510629 (AF047828) - syringomycin synthetase [Pseudomonas syringae pv. syringae]		synthetase	1001
2130	29514403 (20707, 20708)	Novel Protein sim. GBank gj 3515fmb CAA27005] - (X03246) put. DNA ligase (CDC9) (aa 1-755) [Saccharomyces cerevisiae]		synthetase	1009
2131	79607076 (22375, 22376)	Novel Protein sim. GBank gj 3649739 dbj BAA33403] - (A8012226) SecA [Vibrio alginolyticus]		synthetase	1034
2132	20467069 (18047, 18048)	Novel Protein sim. GBank gj 3820584 (AF086791) - carbamoylphosphate synthetase large subunit [Zymomonas mobilis]		synthetase	1010
2133	80026691 (16049, 16050)	Novel Protein sim. GBank gj 3915542 sp O05668 YRSU_MYCLE - HYPOTHETICAL 28.1 KD PROTEIN CB1351.03C		synthetase	1006, 1022
2134	86476300 (11481, 11482)	Novel Protein sim. GBank gj 3915673 sp P45745 DHBF_BACSU - PROBABLE SERINE ACTIVATING ENZYME		synthetase	1029
2135	27980199 (3399, 3400)	Novel Protein sim. GBank gj 3928737 emb CAAA22233] - (A1034355) seryl-tRNA synthetase [Streptomyces coelicolor]		synthetase	1006
2136	20295118 (5039, 5040)	Novel Protein sim. GBank gj 3918474 (AF092918) - MetZ homolog [Pseudomonas alcaligenes]		synthetase	1034
2137	79620871 (13671, 13672)	Novel Protein sim. GBank gj 4062979 dbj BAA36210.1] - (AB017138) epsilon subunit of malonate decarboxylase [Pseudomonas putida]		synthetase	1023

2138	80259151 (15349, 15350)	Novel Protein sim. GBank gi 409397 (L01470) - aminotransferase and dihydrozibotoxine synthase [Bradyrhizobium japonicum]		synthase	1022
2139	30666401 (4589, 4590)	Novel Protein sim. GBank gi 415317dbj BAA039471 - (D16496) cystathione beta-synthase [Saccharomyces cerevisiae]		synthase	1022
2140	27965691 (523, 524)	Novel Protein sim. GBank gi 416912 sp P32469 DPH5_YEAST - DIPHTHINE SYNTHASE (DIPHTAMIDE BIOSYNTHESIS METHYLTRANSFERASE)		synthase	1034
2141	80477734 (13231, 13232)	Novel Protein sim. GBank gi 417442 sp P32483 PABS_STRGR - PARA-AMINOBENZOATE SYNTHASE (P-AMINOBENZOIC ACID SYNTHASE) (PABA SYNTHASE)		synthase	1010, 1012, 1022
2142	30239089 (88867, 88868)	Novel Protein sim. GBank gi 417557 sp P15640 PUR2_ECOLI - PHOSPHORIBOSYLAMINE-GLYCINE LIGASE (GARS) (GLYCINAMIDE RIBONUCLEOTIDE SYNTHETASE) (PHOSPHORIBOSYLGlycinamide SYNTHETASE)		synthase	1044
2143	10173989 (17599, 17600)	Novel Protein sim. GBank gi 4235097 (AF114233) - 5-enopyruvylshikimate 3-phosphate synthase; EPSP synthase [Corynebacterium glutamicum]		synthase	1034
2144	80426873 (15627, 15628)	Novel Protein sim. GBank gi 446728 emb CAB37589 - (AL035569) acetolactate synthase small subunit [Streptomyces coelicolor]		synthase	1017, 1025, 1038
2145	65487401 (16751, 16752)	Novel Protein sim. GBank gi 4481934 emb CAB38518.1 - (AL035640) CDA peptide synthetase I [Streptomyces coelicolor]		synthase	1023

2146	70280512 (12837, 12838)	Novel Protein sim. GBank gi 4539581 emb CAB38500.1 - (AL035636) acetyl-coenzyme A synthetase [Streptomyces coelicolor]		synthase	1010
2147	66377536 (19667, 19668)	Novel Protein sim. GBank gi 4574141 gb AAD23910.1 AF07369 - (AF073698) cysteine synthase [Oryza saliva]		synthase	1010
2148	65487391 (15199, 15200)	Novel Protein sim. GBank gi 477845 pir BA48956 - 4-chlorobenzoate-CoA ligase, 4-CBA-CoA-ligase - Arthrobacter sp.		synthase	1023
2149	98921817 (16243, 16244)	Novel Protein sim. GBank gi 4928924 gb AAD33773.1 AF13910 - (AF139107) glutamyl-tRNA synthetase [Pseudomonas aeruginosa]		synthase	1034
2150	9674778 (14263, 14264)	Novel Protein sim. GBank gi 4980738 gb AAD35331.1 AE00170 - (AE001707) glucose-1-phosphate adenylyltransferase [Thermotoga maritima]		synthase	1044
2151	11398389 (13851, 13852)	Novel Protein sim. GBank gi 4982087 gb AAD36588.1 AE00179 - (AE001799) dihydrodipicolinate synthase [Thermotoga maritima]		synthase	1006
2152	65505067 (19765, 19766)	Novel Protein sim. GBank gi 5019275 emb CAB44431.1 - (AJ132751) xenobiotic/medium-chain fatty acid:CoA ligase form XI-III [Bos taurus]		synthase	1016
2153	87898878 (16715, 16716)	Novel Protein sim. GBank gi 5104430 dbj BAA7945.1 - (AP000060) 306aa long hypothetical spermidine synthase [Aeropyrum pernix]		synthase	1015
2154	20439195 (8743, 8744)	Novel Protein sim. GBank gi 5106357 dbj BAA81639.1 - (AB028629) cystathione beta-synthase [Clostridium perfringens]		synthase	1010

2155	55998116 (7509, 7510)	Novel Protein sim. GBank gil5123663[emb CAB45352.1] - (AL079345) putative porphobilinogen deaminase [Streptomyces coelicolor A3(2)]		synthase	1026
2156	25136896 (17629, 17630)	Novel Protein sim. GBank gil5123663[emb CAB45352.1] - (AL079345) putative porphobilinogen deaminase [Streptomyces coelicolor A3(2)]		synthase	1026
2157	78758460 (11317, 11318)	Novel Protein sim. GBank gil532695 (U13043) - uroporphyrinogen III methyltransferase [Propionibacterium freudenreichii]		synthase	1008, 1021
2158	94678837 (20839, 20840)	Novel Protein sim. GBank gil5420021[emb CAB46413.1] - (AL096743) putative dihydrodipicolinate synthase [Streptomyces coelicolor]		synthase	1022
2159	77575214 (11941, 11942)	Novel Protein sim. GBank gil5421740[dbj BA31358.2] - (AB010284) dihydroorotate [Trypanosoma cruzi]		synthase	1010, 1049
2160	35599540 (10019, 10020)	Novel Protein sim. GBank gil544343ispIP36001[FOLC YEAST -PUTATIVE FOLYLPOYGLUTAMATE SYNTHASE (FOLYLPOY-GAMMA-GLUTAMATE SYNTHETASE) (FPGS)]		synthase	1003, 1016, 1031
2161	94239816 (21987, 21988)	Novel Protein sim. GBank gil5457244[emb CAB46932.1] - (AL096822) putative transmembrane transport protein [Streptomyces coelicolor]		synthase	1003
2162	79595545 (21425, 21426)	Novel Protein sim. GBank gil5457781[emb CAB49368.1] - (AJ248284) PAB0296 [Pyrococcus abyssi]		synthase	1003
2163	10067982 (2321, 2322)	Novel Protein sim. GBank gil5524340[gbf AAD44233.1]AF14377 - (AF143772) PstA [Mycobacterium avium]		synthase	1012
2164	94132731 (14185, 14186)	Novel Protein sim. GBank gil5531404[emb CAB51046.1] - (AJ233397) hypothetical protein [Pseudomonas putida]		synthase	1003

2165	86476194 (5773, 5774)	Novel Protein sim. GBank g 5578859 emb CAB51261.1 - (AL096872) putative cobalamin (5'-phosphate) synthase [Streptomyces coelicolor A3(2)]		synthase	1029, 1030
2166	78182331 (7481, 7482)	Novel Protein sim. GBank g 5578870 emb CAB51278.1 - (AL096872) putative lipoic acid synthetase [Streptomyces coelicolor A3(2)]		synthase	1053
2167	35074144 (10133, 10134)	Novel Protein sim. GBank g 560027 (U15957) - cellulose synthase [Acetobacter xylinum]		synthase	1030
2168	54526955 (14217, 14218)	Novel Protein sim. GBank g 5616226 gb AAD45681.1 AF_5701 - (AF157015) CytH [Streptococcus agalactiae]		synthase	1041
2169	79408325 (10649, 10650)	Novel Protein sim. GBank g 5712094 gb AAD47363.1 AF03857 - (AF038578) 5-enolpyruvylshikimate 3-P synthase [Pseudomonas stutzeri]		synthase	1016
2170	88094823 (13133, 13134)	Novel Protein sim. GBank g 584772 sp P38434 ARGJ_NEIGO - GLUTAMATE N-ACETYLTRANSFERASE (ORNITHINE ACETYLTRANSFERASE) (ORNITHINE TRANSACETYLASE) (OATASE) / AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS)		synthase	1004, 1039
2171	79474879 (8291, 8292)	Novel Protein sim. GBank g 585031 sp P37887 CYSK_BACSU - CYSTEINE SYNTHASE (O-ACETYL SERINE SULFHYDRYLASE) (O-ACETYL SERINE (THIO)-LYASE) (CSASE) (SUPEROXIDE-INDUCIBLE PROTEIN 11) (SOI1)		synthase	1031
2172	16407833 (7153, 7154)	Novel Protein sim. GBank g 625681 pir S45582 - siderophore biosynthesis protein - pbsC - Pseudomonas sp. (strain M114)		synthase	1017

2173	78679022 (8351, 8352)	Novel Protein sim. GBank gil625681 pir S45582 - siderophore biosynthesis protein pbsC - Pseudomonas sp. (strain M14)		synthase	1009
2174	9939518 (11791, 11792)	Novel Protein sim. GBank gil625681 pir S45582 - siderophore biosynthesis protein pbsC - Pseudomonas sp. (strain M14)		synthase	1003
2175	79483289 (13243, 13244)	Novel Protein sim. GBank gil628702 pir S45213 - glutamate-tRNA ligase homolog - Escherichia coli		synthase	1006
2176	87896313 (17213, 17214)	Novel Protein sim. GBank gil642524 (U17699) - glutamyl tRNA synthetase [Azospirillum brasiliense]		synthase	1015
2177	78524800 (14601, 14602)	Novel Protein sim. GBank gil699377 (U15187) - alpha,alpha-trehalose-phosphate synthase [Mycobacterium leprae]		synthase	1008, 1009, 1010, 1023, 1039
2178	80429952 (18481, 18482)	Novel Protein sim. GBank gil728831 sp P39188 ALU1_HUMAN - ALU SUBFAMILY J WARNING ENTRY		synthase	1038
2179	11389090 (22587, 22588)	Novel Protein sim. GBank gil728831 sp P39188 ALU1_HUMAN - ALU SUBFAMILY J WARNING ENTRY		synthase	1022
2180	13088828 (14679, 14680)	Novel Protein sim. GBank gil729295 sp P40109 DAPA_BRELA-DIHYDRODIPICOLINATE SYNTHASE (DHDPSS)		synthase	1024
2181	21395338 (16041, 16042)	Novel Protein sim. GBank gil73041 sp P39823 PSS_BACSU - CDP-DIACYLGLYCEROL-SERINE O-PHOSPHATIDYLTRANSFERASE (PHOSPHATIDYL SERINE SYNTHASE)		synthase	1010

2182	87458361 (17797, 17798)	Novel Protein sim. GBank g 730427 sp P40178 PUR5_STRFR - PUTATIVE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR SYNTHASE)		synthase	1003, 1018
2183	95287762 (8151, 8152)	Novel Protein sim. GBank g 71643 sp P38765 YH19_YEAST - HYPOTHETICAL 32.6 KD PROTEIN IN DAF2-SLT2 INTERGENIC REGION		synthase	1012
2184	10210469 (22171, 22172)	Novel Protein sim. GBank g 83137 pir S19411 - acetolactate synthase homolog - yeast (Saccharomyces cerevisiae)		synthase	1004
2185	79972902 (14987, 14988)	Novel Protein sim. GBank g 837256 (U25130) - syringomycin biosynthesis enzyme [Pseudomonas syringae pv. syringae]		synthase	1003
2186	19542930 (15753, 15754)	Novel Protein sim. GBank g 837256 (U25130) - syringomycin biosynthesis enzyme [Pseudomonas syringae pv. syringae]		synthase	1003
2187	30783871 (2871, 2872)	Novel Protein sim. GBank g 837334 (U26444) - partial ORF with similarity to E. coli EntF, Swiss-Prot Accession Number P11454 [Bacillus subtilis]		synthase	1003
2188	65848012 (12783, 12784)	Novel Protein sim. GBank g 870807 (L43074) - putative [Streptomyces glaucescens]		synthase	1054
2189	46854319 (21873, 21874)	Novel Protein sim. GBank g 870807 (L43074) - putative [Streptomyces glaucescens]		synthase	1029
2190	47656283 (3253, 3254)	Novel Protein sim. GBank g 870809 (L43074) - putative [Streptomyces glaucescens]		synthase	1029
2191	20477474 (19617, 19618)	Novel Protein sim. GBank g 870809 (L43074) - putative [Streptomyces glaucescens]		synthase	1022
2192	20466347 (6939, 6940)	Novel Protein sim. GBank g 974782 emb CAA89019 - (Z49150) cobalamin-independent methionine synthase [Solenostemon scutellarioides]		synthase	1010, 1058

2193	78755742 (21233, 21234)	Novel Protein sim. GBank gj 987116 emb CAA60476 - (X86780) cystathione synthase [Streptomyces hygroscopicus]		synthase	1026, 1030, 1039
2194	10195193 (19583, 19584)	Novel Protein sim. GBank gj 98800 pir S17768 - dehydroquinate synthase (EC 4.6.1.3) - Mycobacterium tuberculosis		synthase	1003
2195	47657159 (1891, 1892)	Novel Protein sim. GBank gj 959652 gb AD34439.1 - (AF135060) fibrillin-2 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	tgf	1029
2196	11816129 (20851, 20852)	Novel Protein sim. GBank gj 959652 gb AD34439.1 - (AF135060) fibrillin-2 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	tgf	1006, 1031
2197	8580363 (1765, 1766)	Novel Protein sim. GBank gj 3327808 (AF051344) - latent transforming growth factor-beta binding protein 4S [Homo sapiens]		tgf	1003, 1006, 1017, 1022, 1023, 1025, 1039, 1049
2198	17903887 (11853, 11854)	Novel Protein sim. GBank gj 1073810 pir A64092 - acetyl coenzyme A acetyltransferase (thiolase) (fadA) homolog - Haemophilus influenzae (strain Rd KW20)		thiolase	1008
2199	20634789 (14355, 14356)	Novel Protein sim. GBank gj 1073810 pir A64092 - acetyl coenzyme A acetyltransferase (thiolase) (fadA) homolog - Haemophilus influenzae (strain Rd KW20)		thiolase	1004
2200	29221188 (21727, 21728)	Novel Protein sim. GBank gj 1073810 pir A64092 - acetyl coenzyme A acetyltransferase (thiolase) (fadA) homolog - Haemophilus influenzae (strain Rd KW20)		thiolase	1030
2201	30274486 (4617, 4618)	Novel Protein sim. GBank gj 112330 pir A39368 - sterol carrier protein 2 precursor - rat		thiolase	1006, 1025, 1026

2202	39507927 (14363, 14364)	Novel Protein sim. GBank g 135753 isp P27796 THIK_YEAST - 3- KETOACYL-COA THIOLASE PEROXISOMAL PRECURSOR (BETA-KETO THIOLASE) (ACETYL-COA ACYLTRANSFERASE) (PEROXISOMAL 3-OXOACYL-COA THIOLASE)		thiolase	1004, 1024
2203	80492026 (9311, 9312)	Novel Protein sim. GBank g 135754 isp P14611 THIL_ALCEU - ACETYL- COA ACETYLTRANSFERASE (ACETOACETYL- COA THIOLASE)		thiolase	1012
2204	85817166 (10339, 10340)	Novel Protein sim. GBank g 1666114 emb [CAB05060] - (Z32098) fadA5 [Mycobacterium tuberculosis]		thiolase	1019
2205	10067193 (3213, 3214)	Novel Protein sim. GBank g 506695 (U10895) - PcaF [Pseudomonas putida]		thiolase	1017
2206	28772814 (9863, 9864)	Novel Protein sim. GBank g 5174389 ref NP_005882.1 pACAT - acetyl- Coenzyme A acetyltransferase 2 (acetoacetyl) Coenzyme A thiolase; CT		thiolase	1026
2207	30265930 (12143, 12144)	Novel Protein sim. GBank g 5174389 ref NP_005882.1 pACAT - acetyl- Coenzyme A acetyltransferase 2 (acetoacetyl) Coenzyme A thiolase; CT		thiolase	1026
2208	29887323 (8315, 8316)	Novel Protein sim. GBank g 124653 emb [CAA64368] - (X94742) olfactory receptor 2 [Gallus gallus]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	tm7	1008
2209	11672531 (5735, 5736)	Novel Protein sim. GBank g 3983382 (AF102527) - olfactory receptor E3 [Mus musculus]	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	tm7	1010
2210	27243212 (12439, 12440)	Novel Protein sim. GBank g 2135416 pir1 38022 - hypothetical protein - human		tm7	1023

2211	5721829 (7615, 7616)	Novel Protein sim. GBank gil3983360 (AF102516) - olfactory receptor A3 [Mus musculus]		tm7	1058
2212	85549686 (22235, 22236)	Novel Protein sim. GBank gil7288381sp P39195 ALU8 HUMAN - !!! ALU SUBFAMILY SX WARNING ENTRY !!!		tm7	1053, 1054
2213	80230431 (22343, 22344)	Novel Protein sim. GBank gil434044 (L26907) - immunoglobulin heavy chain [Homo sapiens]		tnf	1010
2214	79254664 (5455, 5456)	Novel Protein sim. GBank gil1359601sp P25118 TNR1_MOUSE - TUMOR NECROSIS FACTOR RECEPTOR 1 PRECURSOR (P60) (TNF-R1) (P55)	Contains protein domain (PF00020) - TNFR/NGFR cysteine-rich region	tnfreceptor	1003
2215	20292124 (14691, 14692)	Novel Protein sim. GBank gil45062271ref NP_002799.1 pPSMD - proteasome (prosome, macropain) 26S subunit, non-ATPase, 2		tnfreceptor	1034
2216	78077205 (20895, 20896)	Novel Protein sim. GBank gil29833859 (AE000742) - general secretion pathway protein E [Aquifex aeolicus]	Contains protein domain (PF00437) - Bacterial type II secretion system protein	traffic	1029
2217	10372016 (20017, 20018)	Novel Protein sim. GBank gil48091em CAA44309 - (X62452) YCR601 [Saccharomyces cerevisiae]		traffic	1017
2218	27927777 (5873, 5874)	Novel Protein sim. GBank gil1362781 pir [A57291 cytokine inducible nuclear protein C193 - human	Contains protein domain (PF00023) - Ank repeat	transcriptfactor	1001, 1006, 1022, 1024
2219	5260831 (11221, 11222)	Novel Protein sim. GBank gil135077 sp P08153 SM15_YEAST - TRANSCRIPTIONAL FACTOR SW15	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	1058
2220	17941439 (1941, 1942)	Novel Protein sim. GBank gil222472 dbj BA20844 - (AB002388) KIAA0390 [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	1008
2221	80435026 (153, 154)	Novel Protein sim. GBank gil2648674 (AF031955) - kruppel-related zinc finger protein [Mus musculus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	1004, 1023, 1038, 1039
2222	32302670 (18649, 18650)	Novel Protein sim. GBank gil868160 (U27186) - Cys2/His2 zinc finger protein [Rattus norvegicus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	1000

2223	5715335 (20123, 20124)	Novel Protein sim. GBank g 10511186 (U37696) - Ty1 enhancer activator [Saccharomyces cerevisiae]	Contains protein domain (PF00172) - Fungal Zn(2)-Cys(6) binuclear cluster	transcriptifactor	1058
2224	78902290 (21321, 21322)	Novel Protein sim. GBank g 129733 sp P12383 PDR1_YEAST - PLEIOTROPIC DRUG RESISTANCE REGULATOR PROTEIN 1	Contains protein domain (PF00172) - Fungal Zn(2)-Cys(6) binuclear cluster	transcriptifactor	1017, 1022, 1044, 1054
2225	78792738 (19077, 19078)	Novel Protein sim. GBank g 731945 sp P39529 YJU6_YEAST - PUTATIVE 86.7 KD TRANSCRIPTIONAL REGULATORY PROTEIN IN NUC1-NCE1-INTERGENIC REGION	Contains protein domain (PF00172) - Fungal Zn(2)-Cys(6) binuclear cluster	transcriptifactor	1009
2226	20484404 (18011, 18012)	Novel Protein sim. GBank g 134589 sp P22082 SNF2_YEAST - TRANSCRIPTION REGULATORY PROTEIN SNF2 (SWI/SNF COMPLEX COMPONENT SNF2) (REGULATORY PROTEIN SWI2) (REGULATORY PROTEIN GAM1) (TRANSCRIPTION FACTOR TYE3)	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	transcriptifactor	1010
2227	27966999 (3255, 3256)	Novel Protein sim. GBank g 417373 sp P3259 1STH1_YEAST - NUCLEAR PROTEIN STH1/NPS1	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	transcriptifactor	1006
2228	78767100 (17583, 17584)	Novel Protein sim. GBank g 5106572 gb AAD39760.1 AF14394 - (AF143946) transcriptional activator SRCAP [Homo sapiens]	Contains protein domain (PF00176) - SNF2 and others N-terminal domain	transcriptifactor	1003, 1030
2229	71217726 (22905, 22906)	Novel Protein sim. GBank g 1498492 (U63947) - ORF1; homolog to response regulator of a putative signal transduction system [Streptomyces lividans]	Contains protein domain (PF00196) - Bacterial regulatory proteins, luxR family	transcriptifactor	1029, 1053
2230	78765780 (6515, 6516)	Novel Protein sim. GBank g 3367769 emb CAA20094 - (AL031155) putative two-component regulator [Streptomyces coelicolor]	Contains protein domain (PF00196) - Bacterial regulatory proteins, luxR family	transcriptifactor	1017, 1026

2231	86473927 (10911, 10912)	Novel Protein sim. GBank gi 3063697 emb CAA18588.1 - (AL022537) putative myb-protein (partial) [Arabidopsis thaliana]	Contains protein domain (PF00249) - Myb-like DNA- binding domain	transcriptfactor	1054
2232	56649821 (6419, 6420)	Novel Protein sim. GBank gi 1737205 (U75276) - TFIIB related factor hBRF [Homo sapiens]	Contains protein domain (PF00352) - Transcription factor TFIIB repeat	transcriptfactor	1008, 1017, 1023, 1039
2233	55178405 (1295, 1296)	Novel Protein sim. GBank gi 2275218 (AC002337) - G-beta-repeat containing protein isolog [Arabidopsis thaliana]	Contains protein domain (PF00400) - WD domain, G- beta repeat	transcriptfactor	1019
2234	29240889 (11385, 11386)	Novel Protein sim. GBank gi 3367534 (AC004392) - Strong similarity to coatomer alpha subunit (HEPCOP) homolog gi U24105 from Homo sapiens. [Arabidopsis thaliana]	Contains protein domain (PF00400) - WD domain, G- beta repeat	transcriptfactor	1017, 1039, 1044
2235	79605606 (20765, 20766)	Novel Protein sim. GBank gi 588324 sp P38129 T2D4 YEAST - TRANSCRIPTION INITIATION FACTOR TFIID 90 KD SUBUNIT (TAFI-90)	Contains protein domain (PF00400) - WD domain, G- beta repeat	transcriptfactor	1024
2236	80217922 (22397, 22398)	Novel Protein sim. GBank gi 1176014 sp P43609 YFK7 YEAST - HYPOTHETICAL 63.2 KD PROTEIN IN CDC26- SAP155 INTERGENIC REGION	Contains protein domain (PF00569) - Zinc finger present in dystrophin, CBP/p300	transcriptfactor	1003, 1006, 1008, 1010, 1012, 1017, 1022, 1024, 1026, 1031, 1034, 1038, 1044
2237	10254569 (21469, 21470)	Novel Protein sim. GBank gi 2120664 pir S61304 - transcription activator pbrA, iron-regulated - Pseudomonas fluorescens	Contains protein domain (PF00776) - Sigma-70 factor (ECF subfamily)	transcriptfactor	1001
2238	79637451 (20429, 20430)	Novel Protein sim. GBank gi 606340 (U18997) - ORF_o170 [Escherichia coli]	Contains protein domain (PF01272) - Prokaryotic transcription elongation factor GreA/GreB	transcriptfactor	1039
2239	27841808 (4971, 4972)	Novel Protein sim. GBank gi 2114428 (U92702) - Olf-1/EBF-like-2(9L) transcription factor [Mus musculus]	Contains protein domain (PF01833) - IPT/TG domain	transcriptfactor	1022

2240	78938960 (6325, 6326)	Novel Protein sim. GBank gi 114814 sp P14164 BAF1_YEAST - TRANSCRIPTION FACTOR BAF1 (ARS BINDING FACTOR 1) (PROTEIN ABF1) (BIDIRECTIONALLY ACTING FACTOR) (SFB-B) (DNA REPLICATION ENHANCER-BINDING PROTEIN OBF1)		transcriptfactor	1044
2241	78720158 (16207, 16208)	Novel Protein sim. GBank gi 1170943 sp P45128 MFD_HAEIN - TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)		transcriptfactor	1046, 1030
2242	20436004 (15231, 15232)	Novel Protein sim. GBank gi 1172617 sp P44647 PRIA_HAEIN - PRIMOSOMAL PROTEIN N (REPLICATION FACTOR Y)		transcriptfactor	1024
2243	11290953 (7457, 7458)	Novel Protein sim. GBank gi 1174555 sp Q09813 T111_SCHPO - PUTATIVE TRANSCRIPTION INITIATION FACTOR TFIID 111 KD SUBUNIT (TBP-ASSOCIATED FACTOR 111 KD) (TAFII-111)		transcriptfactor	1022
2244	57294417 (12943, 12944)	Novel Protein sim. GBank gi 1176014 sp P43609 YFK7_YEAST - HYPOTHETICAL 63.2 KD PROTEIN IN CDC26-SAP155 INTERGENIC REGION		transcriptfactor	1004, 1006, 1010, 1039, 1044
2245	79775930 (18219, 18220)	Novel Protein sim. GBank gi 131653 sp P25502 PUT3_YEAST - PROLINE UTILIZATION TRANS-ACTIVATOR		transcriptfactor	1003, 1008, 1025
2246	78904093 (14719, 14720)	Novel Protein sim. GBank gi 135633 sp P29055 TF2B_YEAST - TRANSCRIPTION INITIATION FACTOR IIB (TFIIB) (TRANSCRIPTION FACTOR E)		transcriptfactor	1008, 1034, 1039
2247	80093268 (6671, 6672)	Novel Protein sim. GBank gi 1710260 sp Q10607 RHO_MYCTU - TRANSCRIPTION TERMINATION FACTOR RHO HOMOLOG		transcriptfactor	1004, 1010

2248	27965836 (2631, 2632)	Novel Protein sim. GBank gi 171029 sp P52157 RHO_STRL1 - TRANSCRIPTION TERMINATION FACTOR RHO		transcriptfactor	1006
2249	21391828 (1683, 1684)	Novel Protein sim. GBank gi 1723872 sp P53160 YGH4_YEAST - HYPOTHETICAL 11.5 KD PROTEIN IN RPL6- HSF1 INTERGENIC REGION		transcriptfactor	1006
2250	29345186 (1981, 1982)	Novel Protein sim. GBank gi 173056 (M60486) - TSM1 [Saccharomyces cerevisiae]		transcriptfactor	1022
2251	8499361 (10855, 10856)	Novel Protein sim. GBank gi 173056 (M60486) - TSM1 [Saccharomyces cerevisiae]		transcriptfactor	1024
2252	78768290 (11079, 11080)	Novel Protein sim. GBank gi 173056 (M60486) - TSM1 [Saccharomyces cerevisiae]		transcriptfactor	1008
2253	65892357 (8561, 8562)	Novel Protein sim. GBank gi 1730761 sp P53930 YNK7_YEAST - HYPOTHETICAL 26.0 KD PROTEIN IN CYB5- LEU4 INTERGENIC REGION		transcriptfactor	1016
2254	23207496 (10467, 10468)	Novel Protein sim. GBank gi 2114430 (U92703) - Olf-1/EBF-like-3 transcription factor [Mus musculus]		transcriptfactor	1003
2255	11340161 (19391, 19392)	Novel Protein sim. GBank gi 2131783 pir S64953 - hypothetical protein YLR116w - yeast (Saccharomyces cerevisiae)		transcriptfactor	1022
2256	27842471 (5895, 5896)	Novel Protein sim. GBank gi 2132251 pir S65289 - hypothetical protein YPL258c - yeast (Saccharomyces cerevisiae)		transcriptfactor	1003, 1004, 1010, 1022, 1024, 1034, 1054
2257	30635556 (7885, 7886)	Novel Protein sim. GBank gi 12132251 pir S65289 - hypothetical protein YPL258c - yeast (Saccharomyces cerevisiae)		transcriptfactor	1026
2258	29354034 (21885, 21886)	Novel Protein sim. GBank gi 2137676 pir 48694 - probable transcription factor NFE2L1 - mouse (Saccharomyces cerevisiae)		transcriptfactor	1009
2259	80216400 (6367, 6368)	Novel Protein sim. GBank gi 2497197 sp Q05027 YM75_YEAST - HYPOTHETICAL 17.3 KD PROTEIN IN RNA1- RNT1 INTERGENIC REGION		transcriptfactor	1001, 1004, 1006, 1008, 1022, 1024, 1025, 1034, 1039

2260	21641022 (15383, 15384)	Novel Protein sim. GBank gil2978255 dbj BAA25190 - (AB007407) myeloid zinc finger protein-2 [Mus musculus]		transcriptifactor	1034
2261	80389522 (12833, 12834)	Novel Protein sim. GBank gil3327882 dbj BAA31745 - (AB016221) SSL1 [Schizosaccharomyces pombe]		transcriptifactor	1029
2262	80082154 (12965, 12966)	Novel Protein sim. GBank gil3914014 sp P96389 MFD_MYCTU-TRANSCRIPTION-REPAIR_COUPLING FACTOR (TRCF)		transcriptifactor	1024, 1034
2263	27974709 (20761, 20762)	Novel Protein sim. GBank gil4808947 gb AAD30028.1 AF11966 - (AF119667) lipopysis-stimulated remnant receptor alpha subunit [Rattus norvegicus]		transcriptifactor	1004
2264	65474375 (1853, 1854)	Novel Protein sim. GBank gil4850263 emb CAB43032.1 - (AJ239087) hypothetical protein [Streptomyces coelicolor]		transcriptifactor	1038
2265	20728452 (3997, 3998)	Novel Protein sim. GBank gil4982509 gb AAD36969.1 AC000010 - (AC000107) F17F8.15 [Arabidopsis thaliana]		transcriptifactor	1006
2266	20482148 (18181, 18182)	Novel Protein sim. GBank gil5281316 gb AAD41476.1 AF13312 - (AF133124) transcription factor IIIc63 [Homo sapiens]		transcriptifactor	1010
2267	79759623 (6283, 6284)	Novel Protein sim. GBank gil549671 sp P36093 PHD1_YEAST - PUTATIVE TRANSCRIPTION FACTOR PHD1		transcriptifactor	1023, 1031, 1044
2268	17906471 (17673, 17674)	Novel Protein sim. GBank gil551535 (U14134) - transcription factor IIIA [Homo sapiens]		transcriptifactor	1003
2269	94996914 (1705, 1706)	Novel Protein sim. GBank gil730821 sp P38915 SPT8_YEAST - TRANSCRIPTION FACTOR_SPT8		transcriptifactor	1006, 1022
2270	79754888 (20861, 20862)	Novel Protein sim. GBank gil80741 pir S20912 - regulatory protein whiB - Streptomyces coelicolor		transcriptifactor	1006, 1008, 1024, 1027, 1031
2271	79865209 (17855, 17856)			transcriptifactor	1027, 1038

2272	46783948 (10851, 10852)	Novel Protein sim. GBank gj 121749 sp P04906 GTP_RAT - GLUTATHIONE S-TRANSFERASE P (GST-7-7) (CHAIN 7) (CLASS-PI)	Contains protein domain (PF00043) - Glutathione S-transferases.	transferase	1036
2273	80248630 (18667, 18668)	Novel Protein sim. GBank gj 545858 emb CAB50072.1 - (AJ248286) PAB0773 [Pyrococcus abyssi]	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	transferase	1004, 1006, 1009, 1010, 1022, 1023, 1024, 1034, 1044, 1049
2274	20293224 (7463, 7464)	Novel Protein sim. GBank gj 342593 emb CAA11635 - (AJ223904) argF [Neisseria animalis]	Contains protein domain (PF00185) - Aspartate/ornithine carbamoyltransferase	transferase	1034
2275	78464704 (16485, 16486)	Novel Protein sim. GBank gj 523084 gb AAD41041.1 AF11479 - (AF114793) BioA homolog [Vitreoscilla sp. C1]	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate synthase	transferase	1008
2276	80055541 (20785, 20786)	Novel Protein sim. GBank gj 1170780 sp P44930 LGT_HAEIN - PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE	Contains protein domain (PF00303) - Thymidylate synthase	transferase	1004
2277	20466940 (20821, 20822)	Novel Protein sim. GBank gj 320156 emb CAA07073 - (AJ006514) thymidylate synthetase [Vibrio cholerae]	Contains protein domain (PF00303) - Thymidylate synthase	transferase	1010
2278	85801070 (13977, 13978)	Novel Protein sim. GBank gj 205214 emb CAB08137 - (294752) ksgA [Mycobacterium tuberculosis]	Contains protein domain (PF00398) - Ribosomal RNA adenine dimethylases	transferase	1003, 1008, 1025, 1029
2279	25292618 (22727, 22728)	Novel Protein sim. GBank gj 3322607 (AE001213) - serine hydroxymethyltransferase (glyA)	Contains protein domain (PF00464) - Serine hydroxymethyltransferase	transferase	1026
2280	56653981 (12913, 12914)	Novel Protein sim. GBank gj 498124 gb AAD35802.1 (AE00174 - (AE001743) serine hydroxymethyltransferase [Thermotoga maritima])	Contains protein domain (PF00464) - Serine hydroxymethyltransferase	transferase	1008, 1024
2281	37804168 (22119, 22120)	Novel Protein sim. GBank gj 2078350 (U95923) - transaldolase [Solanum tuberosum]	Contains protein domain (PF00923) - Transaldolase	transferase	1058

2282	80077371 (7087, 7088)	Novel Protein sim. GBank g 117292 sp P45830 RFE_MYCLE - PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE	Contains protein domain (PF00953) - Glycosyl transferase	transferase	1006, 1031, 1034
2283	20289282 (2999, 3000)	Novel Protein sim. GBank g 117203 sp P42315 SCOA_BACSU - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE SUBUNIT A (SUCCINYL COA:3-OXOACID COA-TRANSFERASE)(OXCTA)	Contains protein domain (PF01144) - Coenzyme A transferase	transferase	1010
2284	79320692 (5249, 5250)	Novel Protein sim. GBank g 130327 sp P26647 PLSC_ECOLI - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-AGP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)	Contains protein domain (PF01553) - Acyltransferase	transferase	1022
2285	79387490 (4151, 4152)	Novel Protein sim. GBank g 391437 sp O25903 PLSC_HELPY - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-AGP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)	Contains protein domain (PF01553) - Acyltransferase	transferase	1034
2286	10082215 (2429, 2430)	Novel Protein sim. GBank g 1073011 pir B53652 - rhamnosyltransferase 1 chain B - Pseudomonas aeruginosa	Contains protein domain (PF01553) - Acyltransferase	transferase	1017
2287	11076493 (205, 206)	Novel Protein sim. GBank g 1171068 sp P1967 MURA_BACSU - PROBABLE UDP-N-AACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (ENOYL PYRUVATE TRANSFERASE) (UDPN-N-ACETYLGLUCOSAMINE ENOLPYRUVYL TRANSFERASE) (EPT)	Contains protein domain (PF01553) - Acyltransferase	transferase	1010

2288	79829073 (10291, 10292)	Novel Protein sim. GBank g 117152 emb CAA93155 - (269031) Butyryl-CoA:Acetate Coenzyme A transferase [Thermoanaerobacterium thermosaccharolyticum]		transferase	1039
2289	80078972 (3775, 3776)	Novel Protein sim. GBank g 1184993 (147030) - putative uridylyltransferase/uridylyl-removing enzyme; Method: conceptual translation supplied by author		transferase	1034
2290	46571543 (13229, 13230)	Novel Protein sim. GBank g 134585 isp P49586 CTPT_MOUSE - CHOLINEPHOSPHATE CYTIDYLTRANSFERASE (PHOSPHORYLCHOLINE TRANSFERASE) (CT)		transferase	1029
2291	8489232 (5023, 5024)	Novel Protein sim. GBank g 14148 isp P26725 YUR1_YEAST - PROBABLE MANNOSYLTRANSFERASE YUR1		transferase	1022
2292	78897361 (20609, 20610)	Novel Protein sim. GBank g 1588913 pf 2209416L - ORF 11 [Vibrio cholerae]		transferase	1008
2293	55380088 (5759, 5760)	Novel Protein sim. GBank g 170597 sp Q10396 COBT_MYCTU - PUTATIVE NICOTINATE-NUCLEOTIDE--DIMETHYLBENZIMIDAZOLE PHOSPHORIBOSYL-TRANSFERASE (NN:DBI PRT) (N1-ALPHA-PHOSPHORIBOSYLTRANSFERASE)		transferase	1029
2294	33204690 (7489, 7490)	Novel Protein sim. GBank g 1707936 isp P5298 GLGB_SYNY3 - 1,4-ALPHA-GLUCAN BRANCHING ENZYME (GLYCOGEN BRANCHING ENZYME)		transferase	1026

2295 5520)	25269780 (5519, Novel Protein sim. GBank gi 1707970 sp Q10379 GLNE_MYCTU - PROBABLE GLUTAMATE-AMMONIA-LIGASE ADENYLYLTRANSFERASE (GLUTAMINE- SYNTHETASE ADENYLYLTRANSFERASE) (ATASE)		transferase	1008
2296 7656)	Novel Protein sim. GBank gi 1707970 sp Q10379 GLNE_MYCTU - PROBABLE GLUTAMATE-AMMONIA-LIGASE ADENYLYLTRANSFERASE (GLUTAMINE- SYNTHETASE ADENYLYLTRANSFERASE) (ATASE)		transferase	1004, 1006, 1010, 1017, 1022, 1034
2297 17308)	Novel Protein sim. GBank gi 1842055 (U87258) - unknown protein [Acinetobacter sp. ADP1]		transferase	1036
2298 6390)	Novel Protein sim. GBank gi 2117938 pir JC4985 - UTP-glucosidase-1-phosphate uridylyltransferase (EC 2.7.7.9) - Xanthomonas campestris		transferase	1024
2299 1708)	Novel Protein sim. GBank gi 2120998 pir S70682 - glycosyltransferase homolog - Bordetella pertussis		transferase	1001, 1010, 1012, 1022, 1024, 1027
2300 17012)	Novel Protein sim. GBank gi 2120998 pir S70682 - glycosyltransferase homolog - Bordetella pertussis		transferase	1006, 1031, 1034
2301 20420)	Novel Protein sim. GBank gi 2245026 sp emb CAB10446.1 - (Z97341) hypothetical protein [Arabidopsis thaliana]		transferase	1025
2302 9938)	Novel Protein sim. GBank gi 2493000 sp Q09450 SCOT_CAEEL - PROBABLE SUCCINYL-COA:3-KETOACID- COENZYME A TRANSFERASE PRECURSOR (3- OXOACID COA-TRANSFERASE)		transferase	1004
2303 5140)	Novel Protein sim. GBank gi 2495634 sp P76518 YFDE_ECOLI - HYPOTHETICAL 43.3 KD PROTEIN IN EVGS- GLK INTERGENIC REGION		transferase	1001, 1003, 1004, 1006, 1012, 1023, 1024, 1039, 1044

2304	13082122 (10941, 10942)	Novel Protein sim. GBank gi 2995299 emb CAA18328 - (AL022268) putative tRNA delta(2)-isopentenylpyrophosphate transferase [Streptomyces coelicolor]		transferase	1024
2305	79179155 (13469, 13470)	Novel Protein sim. GBank gi 2995299 emb CAA18328 - (AL022268) putative tRNA delta(2)-isopentenylpyrophosphate transferase [Streptomyces coelicolor]		transferase	1027
2306	20291865 (22607, 22608)	Novel Protein sim. GBank gi 3080473 emb CAA18668 - (AL022602) udp-n-acetylglucosamine--n-acetylglucosaminyl-(pentapeptide) pyrophosphoryl-undecaprenol n-acetylglucosamine transferase [Mycobacterium leprae]		transferase	1034
2307	27971087 (9283, 9284)	Novel Protein sim. GBank gi 3122419 spi O06594 NADC_MYCTU - NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING) (QUINOLINATE PHOSPHORIBOSYLTRANSFERASE (DECARBOXYLATING)) (QAPRTASE)		transferase	1022
2308	27971122 (17391, 17392)	Novel Protein sim. GBank gi 3122419 spi O06594 NADC_MYCTU - NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING) (QUINOLINATE PHOSPHORIBOSYLTRANSFERASE (DECARBOXYLATING)) (QAPRTASE)		transferase	1022
2309	65707940 (5105, 5106)	Novel Protein sim. GBank gi 3135688 (AF064070) - putative epimerase/dehydratase Whil [Burkholderia pseudomallei]		transferase	1023
2310	11688579 (19253, 19254)	Novel Protein sim. GBank gi 3191978 emb CAA19377 - (AL0223797) putative protein pil uridyltransferase [Streptomyces coelicolor]		transferase	1031

2311	36825652 (4929, 4930)	Novel Protein sim. GBank gil3249555 (AF010184) - coenzyme A transferase PsecoA [Pseudomonas aeruginosa]		transferase	1054
2312	85809290 (7171, 7172)	Novel Protein sim. GBank gil349067 (L22611) - alginic acid synthesis-related protein [Pseudomonas aeruginosa]		transferase	1048
2313	16841248 (17761, 17762)	Novel Protein sim. GBank gil3550619[emb CAA06172] - (AJ004869) UTP-glucose-1-phosphate uridylyltransferase [Streptococcus pneumoniae]		transferase	1023
2314	12993433 (3901, 3902)	Novel Protein sim. GBank gil3888783[dbj BAA34217.1] - (AB008807) glutathione-dependent dehydroascorbate reductase [Rattus rattus]		transferase	1024
2315	79927185 (13609, 13610)	Novel Protein sim. GBank gil3947714[emb CAA05919] - (AJ003164) initiation factor IF2 [Streptococcus agalactiae]		transferase	1003, 1023
2316	80104723 (7259, 7260)	Novel Protein sim. GBank gil4007785[emb CAA51373.1] - (X72850) 3-oxoadipate CoA-succinyl transferase beta subunit [Sphingomonas sp.]		transferase	1008, 1024, 1027
2317	79418620 (3805, 3806)	Novel Protein sim. GBank gil4012131[sp Q02166 TRPD_ARATH - ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE PRECURSOR]		transferase	1053
2318	80502113 (1289, 1290)	Novel Protein sim. GBank gil4014411[sp P31382 PMT2_YEAST - DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSEYLTRANSFERASE 2]		transferase	1004, 1006, 1008, 1010, 1012, 1022, 1038
2319	78706283 (17401, 17402)	Novel Protein sim. GBank gil4014411[sp P31382 PMT2_YEAST - DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSEYLTRANSFERASE 2]		transferase	1023
2320	94665655 (18417, 18418)	Novel Protein sim. GBank gil421095[pir S30688 - hypothetical protein o246 - Escherichia coli		transferase	1022, 1024, 1034

2321	66727502 (14661, 14662)	Novel Protein sim. GBank gi 4520376 dbj BAA75913.1 - (AB024601) uridylyl transferase [Pseudomonas aeruginosa]		transferase	1049
2322	80082195 (21931, 21932)	Novel Protein sim. GBank gi 462660 sp P33412 MUQ1_YEAST - MUQ1 PROTEIN		transferase	1006, 1034, 1039
2323	19515465 (12015, 12016)	Novel Protein sim. GBank gi 475842 ref NP_004742.1 pGCNT - glucosaminyl (N-acetyl) transferase 3, mucin type		transferase	1025
2324	47652696 (6705, 6706)	Novel Protein sim. GBank gi 5596779 emb CAB51426.1 - (AL096884) putative prolipoprotein diacyl(leglycerol) transferase [Streptomyces coelicolor A3(2)]		transferase	1029
2325	78282976 (16383, 16384)	Novel Protein sim. GBank gi 60380 81 dbj BAA06079 - (D28852) STT3 protein [Saccharomyces cerevisiae]		transferase	1034
2326	80249651 (10295, 10296)	Novel Protein sim. GBank gi 628660 pir S37755 - Adenyl-l-transferase - Escherichia coli		transferase	1022, 1024
2327	79625091 (11661, 11662)	Novel Protein sim. GBank gi 732113 sp P39401 MDOB_ECOLI - PHOSPHOGLYCEROL_TRANSFERASE I (PHOSPHATIDYLGLYCEROL-MEMBRANE- OLIGOSACCHARIDE GLYCEROPHOSPHOTRANSFERASE)		transferase	1001
2328	78781997 (7853, 7854)	Novel Protein sim. GBank gi 100139 dbj BAA70806 - (D64006) nitrate transport protein NrtC [Synechocystis sp.]	Contains protein domain (PF00006) - ABC transporter	transport	1009
2329	47657396 (11357, 11358)	Novel Protein sim. GBank gi 1171910 sp P4505 OPPF_HAEIN - OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN_OPPF	Contains protein domain (PF00005) - ABC transporter	transport	1034
2330	94315225 (4459, 4460)	Novel Protein sim. GBank gi 1171919 sp P4692 OPUA_BACSU - GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN OPUAA	Contains protein domain (PF00005) - ABC transporter	transport	1020, 1029, 1053

2331	80249373 (1439, 1440)	Novel Protein sim. GBank gil1723073 sp Q11040 Y081_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01	Contains protein domain (PF00005) - ABC transporter	transport	1004, 1017, 1022, 1024, 1026, 1034
2332	80209521 (19231, 19232)	Novel Protein sim. GBank gil2314642 gb IAAD08506_1 - (AE000646) ABC transporter, ATP-binding protein [Hil1087] [Helicobacter pylori 26695]	Contains protein domain (PF00005) - ABC transporter	transport	1034
2333	80066533 (7547, 7548)	Novel Protein sim. GBank gil2492595 sp Q53193 Y4TR_RHSN - PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TR	Contains protein domain (PF00005) - ABC transporter	transport	1003, 1006, 1022, 1024, 1039
2334	80237936 (9593, 9594)	Novel Protein sim. GBank gil2635771 emb CAB15264 - (Z99120) similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]	Contains protein domain (PF00005) - ABC transporter	transport	1004, 1006, 1008, 1010, 1022, 1024, 1025, 1029, 1034, 1039, 1053
2335	79830982 (8295, 8296)	Novel Protein sim. GBank gil2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	1023, 1024
2336	80028599 (3217, 3218)	Novel Protein sim. GBank gil2791517 emb CAA16054 - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	Contains protein domain (PF00005) - ABC transporter	transport	1006, 1022
2337	21414073 (6477, 6478)	Novel Protein sim. GBank gil731453 sp P40024 YEM6_YEAST - PROBABLE ATP-DEPENDENT TRANSPORTER YER036C	Contains protein domain (PF00005) - ABC transporter	transport	1034
2338	10109176 (14591, 14592)	Novel Protein sim. GBank gil401473 sp P30864 YAFC_ECOLI - HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN ASPU-MLTD INTERGENIC REGION (ORF304)	Contains protein domain (PF00056) - lactate/malate dehydrogenase	transport	1017
2339	78785158 (11363, 11364)	Novel Protein sim. GBank gil1077355 pir S51392 - hypothetical protein YLR375w - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transport	1022

2340	79552712 (1361, 1362)	Novel Protein sim. GBank g 172769 (M88597) - zinc-finger motifs- codons 160-290 [Saccharomyces cerevisiae]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transport	1038
2341	79652949 (9693, 9694)	Novel Protein sim. GBank g 17061841spl Q10877 CTPB,_MYCTU - CATION-TRANSPORTING P-TYPE ATPASE B	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1008, 1024, 1026
2342	78523967 (14033, 14034)	Novel Protein sim. GBank g 17061841spl Q10877 CTPB,_MYCTU - CATION-TRANSPORTING P-TYPE ATPASE B	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1008, 1026
2343	87913501 (17613, 17614)	Novel Protein sim. GBank g 5019367 emb CAB4420.1 - (ALD078610) putative cation transport system component [Streptomyces coelicolor]	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1015
2344	57147843 (18193, 18194)	Novel Protein sim. GBank g 5866655 spl P37617 ATZN_ECOLI - ZINC-TRANSPORTING ATPASE (ZN(II)-TRANSLOCATING P-TYPE ATPASE)	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1004
2345	56020293 (11779, 11780)	Novel Protein sim. GBank g 1728904 spl P38929 ATC2,_YEAST - CALCIUM-TRANSPORTING ATPASE 2 (VACUOLAR CA2+-ATPASE)	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1009, 1022
2346	94292284 (4979, 4980)	Novel Protein sim. GBank g 1728905 spl P39524 ATC3,_YEAST - PROBABLE CALCIUM-TRANSPORTING ATPASE 3 (ENDOPLASMIC RETICULUM CA2+-ATPASE)	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1026
2347	80229737 (7447, 7448)	Novel Protein sim. GBank g 1728935 spl P38995 ATU2,_YEAST - PROBABLE COPPER-TRANSPORTING ATPASE (CU2+-ATPASE)	Contains protein domain (PF00122) - E1-E2 ATPase	transport	1006, 1022, 1034
2348	20435070 (7859, 7860)	Novel Protein sim. GBank g 415091 emb CAA90593 - (Z50728) putative acetyltransferase [Schizosaccharomyces pombe]	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	transport	1024

2349	80408524 (14675) 14676)	Novel Protein sim. GBank g j731964 sp P40892 YJV8_YEAST - PUTATIVE ACETYLTRANSFERASE IN HXT11-HXT8 INTERGENIC REGION	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	transport	1001, 1006, 1008, 1010, 1012, 1022, 1023, 1024, 1026, 1029, 1030, 1034, 1039
2350	25268334 (20461) 20462)	Novel Protein sim. GBank g j2132389 pir S60949 phosphate transport protein, mitochondrial - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	1003
2351	80586648 (16863) 16864)	Novel Protein sim. GBank g j2132987 pir S609050 probable membrane protein YPL134c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	1003, 1017, 1034, 1044, 1058
2352	95412427 (16333) 16334)	Novel Protein sim. GBank g j418141 sp P32089 TXTP_RAT - TRICARBOXYLATE TRANSPORT PROTEIN PRECURSOR (CITRATE TRANSPORT PROTEIN) (CTP) (TRICARBOXYLATE CARRIER PROTEIN)	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	1053
2353	80257854 (7017) 7018)	Novel Protein sim. GBank g j731875 sp P40464 FLX1_YEAST - MITOCHONDRIAL FAD CARRIER PROTEIN FLX1	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	1024
2354	80426219 (14325) 14326)	Novel Protein sim. GBank g j43607 M57689 - sporulation protein [Bacillus subtilis]	Contains protein domain (PF00158) - Sigma-54 transcription factors	transport	1001, 1006, 1010, 1012, 1025, 1031
2355	25321813 (18915) 18916)	Novel Protein sim. GBank g j117598 sp P44849 Y736_HAEIN - HYPOTHETICAL SODIUM-DEPENDENT TRANSPORTER H10736	Contains protein domain (PF00209) - Sodium:neurotransmitter symporter family	transport	1008
2356	81313170 (21309) 21310)	Novel Protein sim. GBank g j3347922 AF075262 - orphan transporter isoform A12 [Mus musculus]	Contains protein domain (PF00209) - Sodium:neurotransmitter symporter family	transport	1054

2357 9024)	54850773 (9023, Novel Protein sim. GBank gij1113621 sp P08440 ALF_MAIZE - FRUCTOSE- BISPHOSPHATE ALDOLASE, CYTOPLASMIC ISOZYME	Contains protein domain (PF00274) - Fructose- bisphosphate aldolase class-	transport	1054
2358 1612)	9844869 (1611, Novel Protein sim. GBank gij3913016 sp P74309 ALF1_SYNY3- FRUCTOSE-BISPHOSPHATE ALDOLASE CLASS I (FBP ALDOLASE)	Contains protein domain (PF00274) - Fructose- bisphosphate aldolase class-	transport	1023
2359 9676)	80216421 (9675, Novel Protein sim. GBank gij5853901 sp P38552 LEG4_RAT - GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN) (L36LBP)	Contains protein domain (PF00337) - Vertebrate galactoside-binding lectins	transport	1022
2360 16750)	11073361 (16749, Novel Protein sim. GBank gij1172714 sp P42015 PTGA_BACST - PTS SYSTEM, GLUCOSE-SPECIFIC IIABC COMPONENT (IIABC-GLC) (GLUCOSE- PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC COMPONENT) (EII-GLC / EII-GLC)	Contains protein domain (PF00358) - phosphoenolpyruvate- dependent sugar phosphotransferase system, EIIA 1	transport	1034
2361 5192)	27371658 (5191, Novel Protein sim. GBank gij1214621 sp P21345 GLTP_ECOLI - PROTON GLUTAMATE SYMPORT PROTEIN (GLUTAMATE-ASPARTATE CARRIER PROTEIN)	Contains protein domain (PF00375) - Sodium:dicarboxylate symporter family	transport	1022
2362 4040)	27368629 (4039, Novel Protein sim. GBank gij1363095 pir JC4149 - adipocyte amino acid transporter - mouse	Contains protein domain (PF00375) - Sodium:dicarboxylate symporter family	transport	1022
2363 9666)	80193068 (9665, Novel Protein sim. GBank gij2131405 pir S70127 - hypothetical protein YDR267c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00400) - WD domain, G- beta repeat	transport	1058
2364 16374)	25264200 (16373, Novel Protein sim. GBank gij11691271 sp P46840 CTPB_MYCLE - CATION- TRANSPORTING P-TYPE ATPASE B	Contains protein domain (PF00403) - Heavy-metal- associated domain	transport	1039, 1054
2365 3100)	85801231 (3099, Novel Protein sim. GBank gij1706184 sp Q10871 CTPB_MYCTU - CATION- TRANSPORTING P-TYPE ATPASE B	Contains protein domain (PF00403) - Heavy-metal- associated domain	transport	1025, 1026

2366	94666779 (10759, 10760)	Novel Protein sim. GBank gii11703301isp1 P44622 HOFB_HAEIN - PROTEIN TRANSPORT PROTEIN HOFB HOMOLOG	Contains protein domain (PF00437) - Bacterial type II secretion system protein	transport	1004
2367	80058669 (13627, 13628)	Novel Protein sim. GBank gii29600901emb CAA17988.1 - (AL022121) dppA [Mycobacterium tuberculosis]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	transport	1010
2368	11069213 (9953, 9954)	Novel Protein sim. GBank gii51039431dbj BAA19259.1 - (AP0000059) 802aa long hypothetical oligopeptide-binding protein dppA [Aeropyrum pernix]	Contains protein domain (PF00496) - Bacterial extracellular solute-binding proteins, family 5	transport	1034
2369	80053484 (10723, 10724)	Novel Protein sim. GBank gii16526641dbj BAA17584 - (D90907) glutamine- binding periplasmic protein [Synechocystis sp.]	Contains protein domain (PF00528) - Binding-protein- dependent transport systems inner membrane component	transport	1004, 1024, 1026, 1027, 1034
2370	37797007 (11875, 11876)	Novel Protein sim. GBank gii4210905gb AAD12048.1 - (AF045609) AgIG [Sinorhizobium meliloti]	Contains protein domain (PF00528) - Binding-protein- dependent transport systems inner membrane component	transport	1012
2371	79646649 (9343, 9344)	Novel Protein sim. GBank gii1171919isp1 P46920 OPUA_BACSU - GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN OPUA	Contains protein domain (PF00571) - CBS domain	transport	1039
2372	20456475 (21405, 21406)	Novel Protein sim. GBank gii4981710gb AAD36237.1 AE00177 - (AE001773) Mg2+ transporter MgIE, putative [Thermotoga maritima]	Contains protein domain (PF00571) - CBS domain	transport	1010
2373	78780990 (19837, 19838)	Novel Protein sim. GBank gii2492595isp1 Q53193 Y4TR_RHISN - PROBABLE PEPTIDE ABC TRANSPORTER ATP- BINDING PROTEIN Y4TR	Contains protein domain (PF00633) - Helix-hairpin- helix motif.	transport	1022
2374	80066596 (5677, 5678)	Novel Protein sim. GBank gii2833482isp1 Q57124 Y895_HAEIN - HYPOTHETICAL PROTEIN H10895	Contains protein domain (PF00873) - AcrB/AcrD/AcrF family	transport	1010

2375	8761512 (13513, 13514)	Novel Protein sim. GBank gil529013 (U12598) - multidrug resistance protein [Escherichia coli]	Contains protein domain (PF00873) - AcrB/AcrD/AcrF family	transport	1004
2376	78728948 (2505, 2506)	Novel Protein sim. GBank gil1730701isp1P537571YN9A_YEAST - HYPOTHETICAL 37.9 KD PROTEIN IN BIO3-HXT17 INTERGENIC REGION	Contains protein domain (PF01263) - Aldose 1-epimerase	transport	1017, 1023, 1044
2377	79372607 (9929, 9930)	Novel Protein sim. GBank gil3916254 (AF087669) - ExxB [Bordetella bronchiseptica]	Contains protein domain (PF01618) - Mota/TolQ/ExbB proton channel family	transport	1004
2378	79854562 (2375, 2376)	Novel Protein sim. GBank gil1657525[gbf]AAB18053.1 - (U73857) hypothetical protein [Escherichia coli]	Contains protein domain (PF01810) - LysE type translocator	transport	1023, 1024
2379	80254977 (20173, 20174)	Novel Protein sim. GBank gil1001352[dbj]BAA10839 - (D64006) ABC transporter [Synechocystis sp.]		transport	1004
2380	79831854 (14615, 14616)	Novel Protein sim. GBank gil1006620[dbj]BAA10776 - (D64005) ABC transporter [Synechocystis sp.]		transport	1023
2381	20296566 (21829, 21830)	Novel Protein sim. GBank gil1019381[dbj]BAA05389 - (D26358) CmpD [Synechococcus sp.]		transport	1034
2382	9860664 (16733, 16734)	Novel Protein sim. GBank gil1073007[pir][S54001 - pyoverdine synthetase E - Pseudomonas aeruginosa		transport	1008
2383	80020208 (22465, 22466)	Novel Protein sim. GBank gil1073610[pir][S47672 - ugPB protein - Escherichia coli		transport	1006, 1009, 1026
2384	19521018 (11997, 11998)	Novel Protein sim. GBank gil114104[sp1]P08531 ARAG_ECOLI - L-ARABINOSE TRANSPORT ATP-BINDING PROTEIN ARAG		transport	1010
2385	80502738 (14383, 14384)	Novel Protein sim. GBank gil114105[sp1]P08532 ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		transport	1012, 1024

2386	9937297 (18231, 18232)	Novel Protein sim. GBank gi 114116 sp P18275 ARCD_PSEAE - ARGININE/ORNITHINE ANTIPORTER		transport	1017
2387	80243716 (16121, 16122)	Novel Protein sim. GBank gi 114302 sp P13587 ATN1_YEAST - SODIUM TRANSPORT ATPASE_1		transport	1006, 1008, 1022, 1025, 1038
2388	17289360 (12653, 12654)	Novel Protein sim. GBank gi 1149693 emb CAA60220 - (X86499) rbsc [Clostridium perfringens]		transport	1054
2389	79263591 (2665, 2666)	Novel Protein sim. GBank gi 115124 sp P21629 BRAF_PSEAE - HIGH- AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN BRAF		transport	1039
2390	79614254 (5811, 5812)	Novel Protein sim. GBank gi 15125 sp P21630 BRAF_PSEAE - HIGH- AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN BRAF		transport	1039
2391	80422492 (22853, 22854)	Novel Protein sim. GBank gi 15126 sp P25185 BRAZ_PSEAE - BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM III CARRIER PROTEIN (LIV-III)		transport	1010, 1025
2392	13417865 (16315, 16316)	Novel Protein sim. GBank gi 168265 sp P44785 ABC_HAEIN - ATP- BINDING PROTEIN ABC		transport	1022, 1027, 1031
2393	11076821 (461, 462)	Novel Protein sim. GBank gi 1169126 sp P46839 CTPA_MYCLE - CATION- TRANSPORTING P-TYPE ATPASE A		transport	1010
2394	78460685 (19535, 19536)	Novel Protein sim. GBank gi 1169127 sp P46840 CTPB_MYCLE - CATION- TRANSPORTING P-TYPE ATPASE B		transport	1026
2395	80079346 (5479, 5480)	Novel Protein sim. GBank gi 1169168 sp P4508 CYDC_HAEIN - TRANSPORT ATP-BINDING PROTEIN CYDC		transport	1034
2396	18595932 (8453, 8454)	Novel Protein sim. GBank gi 116987 sp P4353 GC20_YEAST - GCN20 PROTEIN		transport	1008, 1044

2397	28987632 (17453, 17454)	Novel Protein sim. GBank g 1169871 sp P43535 GC20_YEAST - GCN20 PROTEIN		transport	1022
2398	71745079 (3465, 3466)	Novel Protein sim. GBank g 1170944 sp P4484 MGLA_HAEIN - GALACTOSIDE TRANSPORT ATP-BINDING PROTEIN MGLA		transport	1008, 1026, 1044
2399	94666547 (21509, 21510)	Novel Protein sim. GBank g 1171909 sp P45052 OPPD_HAEIN - OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPD		transport	1006
2400	20833147 (19275, 19276)	Novel Protein sim. GBank g 1175159 sp P44536 YFGK_HAEIN - HYPOTHETICAL GTP-BINDING PROTEIN H10136		transport	1010
2401	13088523 (1401, 1402)	Novel Protein sim. GBank g 1175856 sp P44742 YEIM_HAEIN - HYPOTHETICAL PROTEIN H10519		transport	1027
2402	17890802 (9351, 9352)	Novel Protein sim. GBank g 1175959 sp P43548 AGP3_YEAST - GENERAL AMINO ACID PERMEASE AGP3		transport	1039
2403	13089849 (10761, 10762)	Novel Protein sim. GBank g 1176299 sp P44993 YIAN_HAEIN - HYPOTHETICAL PROTEIN H11029		transport	1022
2404	37035017 (7693, 7694)	Novel Protein sim. GBank g 1176478 sp P45127 YJK_HAEIN - ABC TRANSPORTER ATP-BINDING PROTEIN H11252		transport	1012
2405	29350615 (4301, 4302)	Novel Protein sim. GBank g 1176917 sp P24088 YSY1_YEAST - HYPOTHETICAL 65.4 KD PROTEIN IN SUBTELOMERIC Y' REPEAT REGION OF CHR XV (ORF 1)		transport	1009
2406	10239143 (9317, 9318)	Novel Protein sim. GBank g 117703 sp P42308 YXIQ_BACSU - HYPOTHETICAL 45.5 KD PROTEIN IN BGLS - KATB INTERGENIC REGION		transport	1039

2407	86464282 (20679, 20680)	Novel Protein sim. GBank g 19898 sp P15031 FECE_ECOLI - IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE		transport	1029
2408	78182035 (21025, 21026)	Novel Protein sim. GBank g 19898 sp P15031 FECE_ECOLI - IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE		transport	1029, 1053
2409	10179605 (16349, 16350)	Novel Protein sim. GBank g 120569 sp P10115 FTSE_ECOLI - CELL DIVISION ATP-BINDING PROTEIN FTSE		transport	1034
2410	79258172 (21903, 21904)	Novel Protein sim. GBank g 1206033 (U48776) - dicarboxylic acid transport protein [Pseudomonas putida]		transport	1001
2411	77520861 (14303, 14304)	Novel Protein sim. GBank g 1215732 (U48718) - OphB [Agrobacterium tumefaciens]		transport	1049
2412	79853104 (4417, 4418)	Novel Protein sim. GBank g 1215733 (U48718) - OphC [Agrobacterium tumefaciens]		transport	1017
2413	80242385 (18707, 18708)	Novel Protein sim. GBank g 123903 sp P23585 HXT2_YEAST - HIGH-AFFINITY GLUCOSE TRANSPORTER HXT2		transport	1006, 1008
2414	30614280 (10625, 10626)	Novel Protein sim. GBank g 126694 sp P15685 MAL6T_YEAST - MALTOSE PERMEASE MAL6T (MALTOSE TRANSPORT PROTEIN MAL6T)		transport	1026
2415	78467995 (4991, 4992)	Novel Protein sim. GBank g 129189 sp P08007 OPPF_SALTY - OLIGOPEPTIDE TRANSPORT ATP-BINDING PROTEIN OPPF		transport	1026
2416	17293060 (7193, 7194)	Novel Protein sim. GBank g 131444 sp P07655 PSTB_ECOLI - PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PSTB		transport	1039

2417	85801119 (3405, 3406)	Novel Protein sim. GBank gi 131498 sp P00550 PTMA_ECOLI - PTS SYSTEM, MANNITOL-SPECIFIC IIABC COMPONENT (EI/ABC-MTL) (MANNITOL-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC COMPONENT) (EI-MTL)		transport	1025
2418	36618699 (13525, 13526)	Novel Protein sim. GBank gi 131498 sp P00550 PTMA_ECOLI - PTS SYSTEM, MANNITOL-SPECIFIC IIABC COMPONENT (EI/ABC-MTL) (MANNITOL-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC COMPONENT) (EI-MTL)		transport	1024, 1053
2419	36995961 (9113, 9114)	Novel Protein sim. GBank gi 131823 sp P16552 RRAFB_ECOLI - RAFFINOSE PERMEASE		transport	1001
2420	27850843 (14197, 14198)	Novel Protein sim. GBank gi 132121 sp P02925 RBSB_ECOLI - D-RIBOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR		transport	1006
2421	57293650 (915, 916)	Novel Protein sim. GBank gi 134277 sp P15303 SC23_YEAST - PROTEIN TRANSPORT PROTEIN SEC23		transport	1004, 1006, 1008, 1022, 1023, 1024, 1034
2422	78929734 (18287, 18288)	Novel Protein sim. GBank gi 134277 sp P15303 SC23_YEAST - PROTEIN TRANSPORT PROTEIN SEC23		transport	1044
2423	78740453 (21415, 21416)	Novel Protein sim. GBank gi 134393 sp P17065 SEC2_YEAST - PROTEIN TRANSPORT PROTEIN SEC2		transport	1022, 1026
2424	79850302 (2433, 2434)	Novel Protein sim. GBank gi 134570 sp P47818 CCC1_YEAST - CCC1 PROTEIN		transport	1054
2425	19728998 (16457, 16458)	Novel Protein sim. GBank gi 1346308 sp P23516 HOXN_ALCEU - HIGH-AFFINITY NICKEL TRANSPORT PROTEIN		transport	1003

2426	87467683 (1985, 1986)	Novel Protein sim. GBank gi 1346563 sp P49308 MOCB_RHIME-PUTATIVE RHIZOPINE-BINDING PROTEIN PRECURSOR		transport	1018
2427	80238805 (21955, 21956)	Novel Protein sim. GBank gi 1351826 sp P48240 MTR3_YEAST - mRNA TRANSPORT REGULATOR MTR3		transport	1001, 1003, 1006, 1017, 1044
2428	32745344 (8203, 8204)	Novel Protein sim. GBank gi 1352126 sp P3649B COMB_STRPN - TRANSPORT PROTEIN COMB		transport	1001
2429	20438842 (22811, 22812)	Novel Protein sim. GBank gi 136748 sp P10905 UGPA_ECOLI - SN-GLYCEROL-3-PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN UGPA		transport	1004
2430	79468460 (19829, 19830)	Novel Protein sim. GBank gi 143607 M57689 - sporulation protein [Bacillus subtilis]		transport	1038
2431	66396746 (20977, 20978)	Novel Protein sim. GBank gi 1477568 U50335 - ORF2 [Mycobacterium smegmatis]		transport	1008
2432	33745765 (5503, 5504)	Novel Protein sim. GBank gi 1532204 U68036 - putative ABC transporter intracellular ATPase subunit BlkdD [Streptomyces coelicolor]		transport	1026
2433	78980560 (17143, 17144)	Novel Protein sim. GBank gi 153368 L10064) - Unknown [Streptomyces cinnamoneensis]		transport	1016
2434	11090287 (15147, 15148)	Novel Protein sim. GBank gi 165299 dbj BAA1791 3 - (D90910) lactose: transport system permease protein LacG [Synechocystis sp.]		transport	1006
2435	13036675 (6443, 6444)	Novel Protein sim. GBank gi 165329 dbj BAA1821 3 - (D90812) ABC Transporter [Synechocystis sp.]		transport	1031
2436	56333298 (6423, 6424)	Novel Protein sim. GBank gi 653584 dbj BAA1849 7 - (D90914) hypothetical protein [Synechocystis sp.]		transport	1046
2437	65878502 (18037, 18038)	Novel Protein sim. GBank gi 1669848 U62055 - CzcD [Bacillus subtilis]		transport	1044

2438	95287841 (181, 182)	Novel Protein sim. GBank g 1673324 dbj BAA10536 - (D64003) hypothetical protein [Synechocystis sp.]		transport	1012, 1023
2439	17899210 (11627, 11628)	Novel Protein sim. GBank g 1703456 sp P54678 ATC1_DICDI - CATION-TRANSPORTING ATPASE PAT1		transport	1038, 1044
2440	79574379 (8583, 8584)	Novel Protein sim. GBank g 1706183 sp Q10876 CTPA_MYCTU - CATION-TRANSPORTING P-TYPE ATPASE A		transport	1031
2441	19883972 (8081, 8082)	Novel Protein sim. GBank g 1708982 sp P53390 MEP3_YEAST - AMMONIUM TRANSPORTER MEP3		transport	1001
2442	19881582 (20255, 20256)	Novel Protein sim. GBank g 1709621 sp P51533 PDRA_YEAST - ATP-DEPENDENT PERMEASE PDR10		transport	1001
2443	79837761 (21923, 21924)	Novel Protein sim. GBank g 1711562 sp Q00947 STP1_YEAST - ZINC FINGER PROTEIN STP1		transport	1006, 1008
2444	78920315 (6875, 6876)	Novel Protein sim. GBank g 1711596 sp P53394 SULX_YEAST - PUTATIVE SULFATE TRANSPORTER YPR003C		transport	1039
2445	80219503 (14789, 14790)	Novel Protein sim. GBank g 1711596 sp P53394 SULX_YEAST - PUTATIVE SULFATE TRANSPORTER YPR003C		transport	1006, 1024, 1034
2446	78903156 (16187, 16188)	Novel Protein sim. GBank g 171263 (M88252) - COT1 protein [Saccharomyces cerevisiae]		transport	1017
2447	78917194 (10827, 10828)	Novel Protein sim. GBank g 172207 (M30583) - proline-specific permease [Saccharomyces cerevisiae]		transport	1003, 1039
2448	78379650 (7419, 7420)	Novel Protein sim. GBank g 172308 sp Q11046 Y089_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.09		transport	1008

2449	71202745 (6797, 6798)	Novel Protein sim. GBank gi 1723082 sp Q11047 Y08A_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50_10		transport	1030, 1049, 1054
2450	20378302 (6861, 6862)	Novel Protein sim. GBank gi 1723082 sp Q11047 Y08A_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50_10		transport	1004
2451	19881824 (14467, 14468)	Novel Protein sim. GBank gi 17236112 sp P54723 YFIG_BACSU - HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN GLVBC_3'REGION		transport	1034
2452	80216380 (4253, 4254)	Novel Protein sim. GBank gi 1722769 (M88557) - zinc-finger motifs- codons 160-290 [Saccharomyces cerevisiae]		transport	1022, 1038
2453	25248272 (3277, 3278)	Novel Protein sim. GBank gi 17306871 sp P53744 BIO5_YEAST - TRANSMEMBRANE REGULATOR OF KAPA/DAPA TRANSPORT		transport	1003
2454	46774534 (3699, 3700)	Novel Protein sim. GBank gi 17306871 sp P53744 BIO5_YEAST - TRANSMEMBRANE REGULATOR OF KAPA/DAPA TRANSPORT		transport	1041
2455	26338264 (13521, 13522)	Novel Protein sim. GBank gi 17306871 sp P53744 BIO5_YEAST - TRANSMEMBRANE REGULATOR OF KAPA/DAPA TRANSPORT		transport	1023
2456	80078071 (12901, 12902)	Novel Protein sim. GBank gi 17306891 sp P53746 FRE4_YEAST - FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 4 PRECURSOR		transport	1008, 1010, 1017, 1022, 1026, 1027, 1044

2457	84453177 (13355, 13356)	Novel Protein sim. GBank gi 1730689 sp P53746 FRE4_YEAST - FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 4 PRECURSOR		transport	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1026, 1027, 1031, 1034, 1038, 1039, 1044, 1049, 1058
2458	65496218 (8011, 8012)	Novel Protein sim. GBank gi 1730693 sp P53750 YN93_YEAST - HYPOTHETICAL 32.8 KD PROTEIN IN BIO3-HXT17 INTERGENIC REGION		transport	1023
2459	80080820 (9233, 9234)	Novel Protein sim. GBank gi 1730694 sp P53751 YN94_YEAST - PUTATIVE 125.2 KD MEMBRANE GLYCOPROTEIN IN BIO3-HXT17 INTERGENIC REGION		transport	1034
2460	82057237 (10769, 10770)	Novel Protein sim. GBank gi 1730699 sp P53756 YN99_YEAST - PROBABLE ATP-DEPENDENT TRANSPORTER YNR070W		transport	1004, 1016, 1022, 1024, 1034, 1038
2461	79107039 (12451, 12452)	Novel Protein sim. GBank gi 1730702 sp P52924 YN9C_YEAST - HYPOTHETICAL 43.9 KD PROTEIN IN HXT17-PAU6 INTERGENIC REGION		transport	1017, 1044
2462	39529805 (8147, 8148)	Novel Protein sim. GBank gi 1730783 sp P53907 YN04_YEAST - HYPOTHETICAL 84.2 KD PROTEIN IN MFA2-MEP2 INTERGENIC REGION		transport	1006, 1008, 1010
2463	88095667 (9595, 9596)	Novel Protein sim. GBank gi 1742117 dbj BAA14856 - (D90767) Lysine-specific proteinase [Escherichia coli]		transport	1025

2464	13089589 (12147, 12148)	Novel Protein sim. GBank gil 174273 dbj BAA15426 - (D90809) Bicyclomycin resistance protein (Sulfonamide resistance protein). [Escherichia coli]		transport	1022
2465	78945532 (11125, 11126)	Novel Protein sim. GBank gil 174340 emb CAB05917 - (Z83327) transport-associated protein [Salmo salar]		transport	1000, 1030, 1037, 1054
2466	80498768 (3049, 3050)	Novel Protein sim. GBank gil 1750127 (U66480) - YncC [Bacillus subtilis]		transport	1001, 1012, 1034
2467	8009280 (14339, 14340)	Novel Protein sim. GBank gil 178324 dbj BAA11726 - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		transport	1022
2468	20622368 (14871, 14872)	Novel Protein sim. GBank gil 178324 dbj BAA11726 - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		transport	1034
2469	95293078 (1637, 1638)	Novel Protein sim. GBank gil 188135 dbj BAA19377 - (AB001488) PROBABLE TRANSPORT PROTEIN, SIMILAR TO ANTIBIOTIC TRANSPORT-ASSOCIATED PROTEIN ACTI I IN STREPTOMYCES COELICOLOR. [Bacillus subtilis]		transport	1001, 1004, 1006, 1010, 1038
2470	79179287 (14237, 14238)	Novel Protein sim. GBank gil 19196308 (U91632) - sugar transporter [Agrobacterium tumefaciens]		transport	1024
2471	95287780 (12177, 12178)	Novel Protein sim. GBank gil 194571 emb CAB08012 - (Z94043) hypothetical protein [Bacillus subtilis]		transport	1006, 1009, 1010, 1012, 1022, 1034
2472	78386300 (14277, 14278)	Novel Protein sim. GBank gil 2072782 (U62966) - Na+/nucleotide cotransporter [Homo sapiens]		transport	1026
2473	79116464 (2663, 2664)	Novel Protein sim. GBank gil 2114423 (U92466) - osmoregulated proline transporter [Bacillus subtilis]		transport	1013

2474	8724174 (6531, 6532)	Novel Protein sim. GBank g 2119777 pir I40452 - ABC-type transport system probable ATP binding subunit - <i>Bacillus subtilis</i>		transport	1004
2475	25257262 (4175, 4176)	Novel Protein sim. GBank g 2121067 pir I76773 - ATP binding component of ABC-transporter - <i>Escherichia coli</i>		transport	1026
2476	79485405 (9919, 9920)	Novel Protein sim. GBank g 2127458 pir S72583 - abc1 protein - <i>Mycobacterium leprae</i>		transport	1022
2477	30575499 (2685, 2686)	Novel Protein sim. GBank g 2131470 pir S69688 - hypothetical protein YDR406w - yeast (<i>Saccharomyces cerevisiae</i>)		transport	1006, 1008, 1009, 1022, 1024, 1034
2478	80503055 (13665, 13666)	Novel Protein sim. GBank g 2132083 pir S67089 - hypothetical protein YOR197w - yeast (<i>Saccharomyces cerevisiae</i>)		transport	1012, 1023, 1025
2479	94135009 (14745, 14746)	Novel Protein sim. GBank g 2144024 pir S70632 - zinc transporter ZnT-2 - rat		transport	1010
2480	20602749 (15879, 15880)	Novel Protein sim. GBank g 2144024 pir S70632 - zinc transporter ZnT-2 - rat		transport	1004
2481	27977159 (15423, 15424)	Novel Protein sim. GBank g 2150110 (AF001333) - ATP-binding protein [<i>Synechococcus PCC7942</i>]		transport	1022
2482	17708361 (21893, 21894)	Novel Protein sim. GBank g 2160185 (AC000132) - Similar to <i>S. pombe</i> ISP4 (gb D33992). [Arabidopsis thaliana]		transport	1017
2483	80257820 (2647, 2648)	Novel Protein sim. GBank g 225306 lembi CAB10722 - (Z97559) ansP [<i>Mycobacterium tuberculosis</i>]		transport	1003, 1006, 1054
2484	80081984 (1603, 1604)	Novel Protein sim. GBank g 2293416 (AF007800) - MtG [<i>Pseudomonas fluorescens</i>]		transport	1003, 1022
2485	27845441 (22205, 22206)	Novel Protein sim. GBank g 2293448 (AF008930) - transmembrane protein [<i>Bacillus subtilis</i>]		transport	1004

2486 21338)	87452044 (21337, Novel Protein sim. GBank gi 2316016 (U92650) - MRP-like ABC transporter [Arabidopsis thaliana]		transport	1025
2487 5116)	80427306 (5115, Novel Protein sim. GBank gi 2317847 (U44917) - ATP-binding protein [Babesia bovis]		transport	1001, 1006, 1008, 1009, 1022, 1023, 1025, 1027, 1031, 1034, 1038, 1044
2488 12430)	85816902 (12429, Novel Protein sim. GBank gi 2326691 emb CAB11002 - (Z98271) cell division atp-binding protein [Mycobacterium leprae]		transport	1019
2489 2354)	78473414 (2353, Novel Protein sim. GBank gi 2340070 gb AAB67294.1 - (U59485) AttA1 [Agrobacterium tumefaciens]		transport	1026
2490 15464)	10103804 (15463, Novel Protein sim. GBank gi 2333906 (AF008187) - phosphate transport protein [Sinorhizobium melliloti]		transport	1017
2491 8836)	54994364 (8835, Novel Protein sim. GBank gi 2443286 dbj BA22327 - (D86417) YfM		transport	1029
2492 19464)	7945533 (19463, Novel Protein sim. GBank gi 2492533 sp Q46065 AROP_CORG1 - AROMATIC AMINO ACID TRANSPORT PROTEIN AROP (GENERAL AROMATIC AMINO ACID PERMEASE)		transport	1023, 1024, 1025, 1031
2493 790)	13528620 (789, Novel Protein sim. GBank gi 2492558 sp Q47538 TAUB_ECOLI - TAURINE TRANSPORT ATP-BINDING PROTEIN TAUB		transport	1024
2494 9898)	30788857 (9897, Novel Protein sim. GBank gi 2492558 sp Q47538 TAUB_ECOLI - TAURINE TRANSPORT ATP-BINDING PROTEIN TAUB		transport	1003

2495	21636174 (21513, 21514)	Novel Protein sim. GBank gii2492566 sp Q57335 Y036_HAEIN - HYPOTHETICAL ABC TRANSPORTER ATP - BINDING PROTEIN HI0036	transport	1010
2496	866881118 (19339, 19340)	Novel Protein sim. GBank gii2492592 sp P55570 Y4MK_RHISN - PROBABLE ABC TRANSPORTER ATP-BINDING PROTEIN Y4MK	transport	1007
2497	80216476 (17841, 17842)	Novel Protein sim. GBank gii2492605 sp Q08234 YOH5_YEAST - PROBABLE ATP-DEPENDENT TRANSPORTER YOL075C	transport	1003, 1006, 1008, 1010, 1022, 1024, 1034, 1039, 1044
2498	332272358 (7097, 7098)	Novel Protein sim. GBank gii2493015 sp Q64430 AT7A_MOUSE - COPPER - TRANSPORTING ATPASE 1(COPPER PUMP 1)	transport	1000
2499	29254438 (15227, 15228)	Novel Protein sim. GBank gii2495575 sp P75757 YBGR_ECOLI - HYPOTHETICAL 34.7 KD PROTEIN IN PNUC - AROG INTERGENIC REGION	transport	1017
2500	24124816 (10451, 10452)	Novel Protein sim. GBank gii2496701 sp P55552 Y4LL_RHISN - HYPOTHETICAL 91.8 KD PROTEIN Y4LL	transport	1009
2501	78460657 (15151, 15152)	Novel Protein sim. GBank gii2501391 sp Q57130 YE71_HAEIN - HYPOTHETICAL ABC TRANSPORTER PERMEASE PROTEIN HI1471	transport	1026
2502	78375770 (33, 34)	Novel Protein sim. GBank gii2501669 sp P75892 YCDG_ECOLI - HYPOTHETICAL 48.1 KD PROTEIN IN WRBA - PUTA INTERGENIC REGION	transport	1044
2503	20724558 (2639, 2640)	Novel Protein sim. GBank gii2506112 sp P43672 UUP_ECOLI - ABC TRANSPORTER ATP-BINDING PROTEIN UUP	transport	1006

2504	11072654 (21039) 21040)	Novel Protein sim. GBank gi 25061113isp P52648 ZNUC_ECOLI - HIGH-AFFINITY ZINC UPTAKE SYSTEM ATP-BINDING PROTEIN ZNUC		transport	1034
2505	30379550 (17863, 17864)	Novel Protein sim. GBank gi 25061119 isp Q05360 WHIT_LUCCU - WHITE PROTEIN		transport	1026
2506	80225324 (3627, 3628)	Novel Protein sim. GBank gi 2507365 isp P37902 YBEJ_ECOLI - AMINO-ACID ABC TRANSPORTER BINDING PROTEIN IN GLTJ-CUTE INTERGENIC REGION PRECURSOR		transport	1022
2507	17930351 (19649, 19650)	Novel Protein sim. GBank gi 2507365 isp P37902 YBEJ_ECOLI - AMINO-ACID ABC TRANSPORTER BINDING PROTEIN IN GLTJ-CUTE INTERGENIC REGION PRECURSOR		transport	1006, 1038
2508	66727272 (7545, 7546)	Novel Protein sim. GBank gi 2507487 isp P30862 ARTM_ECOLI - ARGinine TRANSPORT SYSTEM PERMEASE PROTEIN ARTM		transport	1049
2509	80417400 (5015, 5016)	Novel Protein sim. GBank gi 254708 dbj BAA22363 - (D89963) Integral membrane protein A [Enterobacter cloacae]		transport	1025
2510	10341484 (4939, 4940)	Novel Protein sim. GBank gi 2578388 emb CAA15479 - (AL008609) ABC-transporter atp-binding protein [Mycobacterium leprae]		transport	1038
2511	78375825 (18881, 18882)	Novel Protein sim. GBank gi 2621685 (AE000842) - ABC transporter [Methanobacterium thermoautotrophicum]		transport	1044
2512	65900577 (671, 672)	Novel Protein sim. GBank gi 2622773 (AE000923) - ABC transporter [Methanobacterium thermoautotrophicum]		transport	1006, 1016
2513	23307657 (7569, 7570)	Novel Protein sim. GBank gi 2632021 emb CAA05580 - (AJ002571) YktD [Bacillus subtilis]		transport	1026

2514	10090583 (1063, 1064)	Novel Protein sim. GBank gi 2633808 emb CAB13310 - (Z99111) similar to hypothetical proteins [Bacillus subtilis]		transport	1017
2515	80498610 (2789, 2790)	Novel Protein sim. GBank gi 2849101 (AE001001) - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]		transport	1004, 1010, 1012, 1024
2516	79168645 (15903, 15904)	Novel Protein sim. GBank gi 2649101 (AE001001) - ABC transporter, ATP-binding protein [Archaeoglobus fulgidus]		transport	1031
2517	20139753 (623, 624)	Novel Protein sim. GBank gi 2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]		transport	1024
2518	20721276 (3839, 3840)	Novel Protein sim. GBank gi 2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]		transport	1006
2519	38920364 (22325, 22326)	Novel Protein sim. GBank gi 2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]		transport	1008
2520	27847422 (6047, 6048)	Novel Protein sim. GBank gi 2656137 (AF015304) - equilibrative nitrobenzylthioinosine-sensitive nucleoside transporter [Rattus norvegicus]		transport	1006
2521	27832757 (21985, 21986)	Novel Protein sim. GBank gi 2661679 emb CAA15784 - (AL009199) probable transmembrane transport protein [Streptomyces coelicolor]		transport	1004
2522	29354697 (13319, 13320)	Novel Protein sim. GBank gi 2663961 sp P30606 TR2_YEAST - MYO-INOSITOL TRANSPORTER 2		transport	1009
2523	79615913 (19383, 19384)	Novel Protein sim. GBank gi 2668595 (AF012537) - arginine and ornithine binding protein [Pseudomonas aeruginosa]		transport	1003, 1016
2524	34408474 (19815, 19816)	Novel Protein sim. GBank gi 2738774 (AF016580) - HufC [Vibrio cholerae]		transport	1030

2525	80056657 (13897, 13898)	Novel Protein sim. GBank gi 2791407 emb CAA16001 - (AL021184) hypothetical protein Rv1473 [Mycobacterium tuberculosis]	transport	1001, 1004, 1008, 1017, 1034, 1044
2526	24114061 (12293, 12294)	Novel Protein sim. GBank gi 2791517 emb CAA16054 - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	transport	1044
2527	79462157 (19737, 19738)	Novel Protein sim. GBank gi 2791517 emb CAA16054 - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	transport	1031
2528	78182864 (20089, 20090)	Novel Protein sim. GBank gi 2791517 emb CAA16054 - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	transport	1053
2529	78459671 (363, 364)	Novel Protein sim. GBank gi 2791639 emb CAA16126.1 - (AL021287) hypothetical protein Rv3041c [Mycobacterium tuberculosis]	transport	1026, 1029, 1053
2530	86472853 (15513, 15514)	Novel Protein sim. GBank gi 2791639 emb CAA16126.1 - (AL021287) hypothetical protein Rv3041c [Mycobacterium tuberculosis]	transport	1000, 1008, 1029
2531	46894672 (14265, 14266)	Novel Protein sim. GBank gi 2801417 (AF007777) - KpsD [Escherichia coli]	transport	1019, 1026
2532	80234832 (15445, 15446)	Novel Protein sim. GBank gi 2808789 emb CAA16218 - (AL021411) putative cobalt transport protein [Streptomyces coelicolor]	transport	1022, 1034
2533	98244110 (22489, 22490)	Novel Protein sim. GBank gi 2828822 (AF012836) - inner membrane protein MalG [Thermococcus litoralis]	transport	1008, 1012, 1026
2534	27981470 (21657, 21658)	Novel Protein sim. GBank gi 2829579 sp P71886 YOF8_MYCTU - HYPOTHETICAL ABC TRANSPORTER.ATP. BINDING PROTEIN CY3G12.08	transport	1006

2535	80026665 (13255, 13256)	Novel Protein sim. GBank g 2829633 spp P75831YBJZ_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP - BINDING PROTEIN YBJZ		transport	1006, 1010, 1024, 1034
2536	79199571 (13995, 13996)	Novel Protein sim. GBank g 2829633 spp P75831YBJZ_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP - BINDING PROTEIN YBJZ		transport	1024
2537	80061336 (1161, 1162)	Novel Protein sim. GBank g 2851440 spp P27854 AARF_ECOLI - UBIQUINONE BIOSYNTHESIS PROTEIN AARF		transport	1004, 1010
2538	27981911 (16311, 16312)	Novel Protein sim. GBank g 2896777 emb CAA17254 - (AL021899) hypothetical protein Rv2040c [Mycobacterium tuberculosis]		transport	1006
2539	27982009 (22777, 22778)	Novel Protein sim. GBank g 2896777 emb CAA17254 - (AL021899) hypothetical protein Rv2040c [Mycobacterium tuberculosis]		transport	1006
2540	10048874 (99, 100)	Novel Protein sim. GBank g 290513 (L10328) - f470 [Escherichia coli]		transport	1039
2541	78932488 (11465, 11466)	Novel Protein sim. GBank g 290592 emb CAA17316 - (AL021928) hypothetical protein Rv0194 [Mycobacterium tuberculosis]		transport	1009, 1039
2542	80248804 (3777, 3778)	Novel Protein sim. GBank g 2916947 emb CAA17555 - (AL021999) hypothetical protein Rv0986 [Mycobacterium tuberculosis]		transport	1004, 1006, 1010, 1017, 1022, 1024, 1034, 1038, 1044
2543	11012248 (727, 728)	Novel Protein sim. GBank g 2960088 emb CAA17986.1 - (AL022121) dppC [Mycobacterium tuberculosis]		transport	1027
2544	25144653 (19899, 19900)	Novel Protein sim. GBank g 2960180 emb CAA18078.1 - (AL022121) proZ [Mycobacterium tuberculosis]		transport	1026

2545	87453369 (16647, 16648)	Novel Protein sim. GBank g 298064 lemb CAA72120 - (Y11250) multi resistance protein [Arabidopsis thaliana]		transport	1025
2546	34415184 (17889, 17890)	Novel Protein sim. GBank g 2982567 lemb CAA05793 - (A 003004) ABC transporter [Rattus norvegicus]		transport	1050
2547	33269118 (11259, 11260)	Novel Protein sim. GBank g 3004482 lemb CAA71076 - (Y09945) putative integral membrane transport protein [Rattus norvegicus]		transport	1000
2548	78402618 (6793, 6794)	Novel Protein sim. GBank g 3024669 spl Q0377 ISSY1_YEAST - PUTATIVE AMINO-ACID PERMEASE SSY1		transport	1009
2549	3373592 (9963, 9964)	Novel Protein sim. GBank g 3025027 spl P75790 YBIT_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBIT		transport	1012
2550	57140383 (18199, 18200)	Novel Protein sim. GBank g 3111095 L16955 - Mdi2p [Saccharomyces cerevisiae]		transport	1010
2551	9736571 (6841, 6842)	Novel Protein sim. GBank g 3183458 spl P75796 YLI_A_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLI_A		transport	1039
2552	21635575 (16797, 16798)	Novel Protein sim. GBank g 3183458 spl P75796 YLI_A_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLI_A		transport	1012, 1058
2553	24122278 (20377, 20378)	Novel Protein sim. GBank g 3183458 spl P75796 YLI_A_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLI_A		transport	1003, 1038
2554	25257745 (12337, 12338)	Novel Protein sim. GBank g 3183561 spl P80866 V296_BACSU - VEGETATIVE PROTEIN 296 (VEG296)		transport	1008

2555	49847260 (19921, 19922)	Novel Protein sim. GBank gii 3258287 dbj BAA30970_1 - (AP0000007) 318aa long hypothetical ATP-binding transport protein [Pyrococcus horikoshii]		transport	1036
2556	25257282 (7961, 7962)	Novel Protein sim. GBank gii 3322404 (AE001199) - K+ transport protein (ntpJ) [Treponema pallidum]		transport	1026
2557	16845211 (7317, 7318)	Novel Protein sim. GBank gii 3341462 emb CAA11688 - (AJ223959) very-long-chain acyl-CoA synthetase related protein [Mus musculus]		transport	1039
2558	80049467 (13931, 13932)	Novel Protein sim. GBank gii 3341854 (AF077856) - PofA [Actinobacillus actinomycetemcomitans]		transport	1010, 1024
2559	78908006 (15581, 15582)	Novel Protein sim. GBank gii 3387977 (AF070598) - ABC transporter [Homo sapiens]		transport	1039
2560	47658242 (16883, 16884)	Novel Protein sim. GBank gii 3414726 (AF047693) - multidrug resistance efflux pump homolog PmrB [Pseudomonas aeruginosa]		transport	1029
2561	80048647 (16157, 16158)	Novel Protein sim. GBank gii 3420605 gb AAC31905_1 - (AF075709) putative FMNH2-dependent monooxygenase [Pseudomonas putida]		transport	1034
2562	78518619 (20021, 20022)	Novel Protein sim. GBank gii 3426107 emb CAA07722 - (AJ007906) FhuC [Rhizobium leguminosarum]		transport	1008, 1026
2563	78470594 (7351, 7352)	Novel Protein sim. GBank gii 3790619 (AF097414) - Efmb [Enterococcus faecium]		transport	1026
2564	17937775 (13011, 13012)	Novel Protein sim. GBank gii 3879893 emb CAA16511 - (AL021571) predicted using Genefinder [Caenorhabditis elegans]		transport	1008
2565	14206971 (8567, 8568)	Novel Protein sim. GBank gii 3913601 sp O2575 EXB1_HELPPY - PUTATIVE BIOPOLYMER TRANSPORT EXBB-LIKE PROTEIN 1		transport	1034

2566	28776366 (16771, 16772)	Novel Protein sim. GBank gi 3914190 sp O35913 OATB_RAT - SODIUM-INDEPENDENT ORGANIC ANION TRANSPORTER 2 (BRAIN DIGOXIN CARRIER PROTEIN) (BRAIN-SPECIFIC ORGANIC ANION TRANSPORTER) (OATP-B1)		transport	1008
2567	25154207 (17203, 17204)	Novel Protein sim. GBank gi 3915465 sp P77254 YFGK_ECOLI - HYPOTHETICAL GTP-BINDING PROTEIN IN XSEA-HISS INTERGENIC REGION		transport	1003
2568	56164940 (10227, 10228)	Novel Protein sim. GBank gi 3915564 sp O32154 YURM_BACSU - HYPOTHETICAL ABC TRANSPORTER PERMEASE PROTEIN YURM		transport	1030
2569	80483336 (21031, 21032)	Novel Protein sim. GBank gi 3915974 sp P75957 YCFV_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YCFV		transport	1012, 1024
2570	46854384 (3509, 3510)	Novel Protein sim. GBank gi 3928723 emb CAA2222 9] - (AL_034355) putative ABC transporter [Streptomyces coelicolor]		transport	1022, 1029
2571	17895529 (20909, 20910)	Novel Protein sim. GBank gi 4098078 U73183 - ABC transporter ATP binding subunit [Streptococcus mutans]		transport	1024
2572	39523886 (1113, 1114)	Novel Protein sim. GBank gi 4106600 emb CAA21355] - (AL031866) ORF32, len.: 236 aa, hypothetical ABC transporter, similar to many, eg: E. coli YBBA_ECOLI (228 aa), 44.0% identity in 225 aa overlap, Fasta scores: opt: 611, E(): 0 [Yersinia pestis]		transport	1004
2573	5603411 (15235, 15236)	Novel Protein sim. GBank gi 4115377 AC005967 - unknown protein [Arabidopsis thaliana]		transport	1058

2574	30190862 (8485, 8486)	Novel Protein sim. GBank gj 4115627 dbj BAA36465j - (AB015053) outer membrane protein (CMF) [Pseudomonas fluorescens]		transport	1008
2575	80029416 (8769, 8770)	Novel Protein sim. GBank gj 41440lembi CAA29254j - (X05810) fhuC product (AA 1-265) [Escherichia coli]		transport	1004, 1022, 1039
2576	81798908 (3993, 3994)	Novel Protein sim. GBank gj 4166581sp P32448 ASF2_YEAST - ANTI-SILENCING PROTEIN 2		transport	1001, 1003, 1004, 1006, 1008, 1009, 1010, 1012, 1017, 1022, 1023, 1024, 1026, 1027, 1031, 1034, 1038, 1039, 1044, 1054
2577	27966079 (13903, 13904)	Novel Protein sim. GBank gj 4174621sp P32319 PEP1_YEAST - VACUOLAR PROTEIN SORTING TARGETING PROTEIN PEP1 PRECURSOR (VACUOLAR CARBOXYPEPTIDASE SORTING RECEPTOR VPS10), (CARBOXYPEPTIDASE Y RECEPTOR) (CPY RECEPTOR)		transport	1034
2578	87447968 (14545, 14546)	Novel Protein sim. GBank gj 4406780 gbj AAD20090j - (AC006532) putative multispanning membrane protein [Arabidopsis thaliana]		transport	1025
2579	80248383 (12823, 12824)	Novel Protein sim. GBank gj 4467368 lembi CAB37661j - (A 224479) cdd4-like protein [Bacillus sphaericus]		transport	1006, 1017, 1024, 1034, 1036
2580	20457555 (3335, 3336)	Novel Protein sim. GBank gj 4514346 dbj BAA75384.1j - (AB013374) MsmX [Bacillus halodurans]		transport	1010
2581	80217915 (21141, 21142)	Novel Protein sim. GBank gj 453646 (L11994) - permease [Saccharomyces cerevisiae]		transport	1004, 1006, 1008, 1022, 1024, 1034

2582	27850298 (2475, 2476)	Novel Protein sim. GBank gi 4584481 emb CAB40684.1 - (AL049587) putative L-asparagine permease [Streptomyces coelicolor]		transport	1010
2583	80077293 (17577, 17578)	Novel Protein sim. GBank gi 4587878 gb AAD25904.1 AF06930 - [Pedicoccus pentosaceus]		transport	1006, 1008, 1034
2584	81871503 (22357, 22358)	Novel Protein sim. GBank gi 4589556 gb BA76800.1 - (AB023173) KIA-A0956 protein [Homo sapiens]		transport	1054
2585	10107856 (4633, 4634)	Novel Protein sim. GBank gi 462712 sp P33592 NIKC_ECOLI - NICKEL TRANSPORT SYSTEM PERMEASE PROTEIN NIKC		transport	1012
2586	80216281 (19743, 19744)	Novel Protein sim. GBank gi 464819 sp P23302 PDR5 YEAST - SUPPRESSOR OF TOXICITY OF SPORIDESMIN		transport	1006, 1022, 1023, 1034
2587	30254796 (21345, 21346)	Novel Protein sim. GBank gi 4678332 emb CAB41143.1 - (AL049658) putative peptide transporter [Arabidopsis thaliana]		transport	1026
2588	30814437 (6433, 6434)	Novel Protein sim. GBank gi 4757573 gb AAD59086.1 AF08287 - [Listeria enterocitica]		transport	1044
2589	20424446 (12675, 12676)	Novel Protein sim. GBank gi 4835822 gb AAD30266.1 AF10217 - (AF102174) glycine betaine transporter BetL [Listeria monocytogenes]		transport	1022
2590	80501478 (16293, 16294)	Novel Protein sim. GBank gi 4835822 gb AAD30266.1 AF10217 - (AF102174) glycine betaine transporter BetL [Listeria monocytogenes]		transport	1006, 1010, 1012

2591	80050373 (3889, 3890)	Novel Protein sim. GBank gi 4877782 gb AAD31429.1 AF12695 - (AF126932) capsule transport protein KpsT [Escherichia coli]		transport	1006, 1039
2592	46741384 (9177, 9178)	Novel Protein sim. GBank gi 4927475 gb AAD33131.1 AF09901 - (AF09901.5) Fe uptake system permease [Streptomyces coelicolor]		transport	1054
2593	20610291 (14235, 14236)	Novel Protein sim. GBank gi 4980606 gb AAD35209.1 AE00169 - (AE00169.7) sugar ABC transporter, ATP-binding protein [Thermotoga maritima]		transport	1006
2594	79871296 (19883, 19884)	Novel Protein sim. GBank gi 4980606 gb AAD35209.1 AE00169 - (AE00169.7) sugar ABC transporter, ATP-binding protein [Thermotoga maritima]		transport	1044
2595	13518389 (8773, 8774)	Novel Protein sim. GBank gi 4980892 gb AAD35474.1 AE00171 - (AE00171.8) ABC transporter, ATP-binding protein [Thermotoga maritima]		transport	1024
2596	71091934 (5327, 5328)	Novel Protein sim. GBank gi 4981115 gb AAD35681.1 AE00173 - (AE00173.4) sugar ABC transporter, permease protein [Thermotoga maritima]		transport	1029, 1053
2597	80501707 (3127, 3128)	Novel Protein sim. GBank gi 5051428 emb CAB4501.1 - (AJ242839) putative ATP-binding protein [Neisseria gonorrhoeae]		transport	1012
2598	78798916 (22827, 22828)	Novel Protein sim. GBank gi 5092051 emb CAA52970 - (X75076) putative amino acid permease 1 [Saccharomyces cerevisiae]		transport	1009, 1023, 1044
2599	32153112 (13103, 13104)	Novel Protein sim. GBank gi 5106228 dbj BA81539.1 - (AP000064) 332aa long hypothetical high-affinity branched-chain amino acid transport permease protein [Aeropyrum pernix]		transport	1039

2600	79636021 (4091, 4092)	Novel Protein sim. GBank gil5456699 gb AAD43625.1 AF07141 - (AF071410) ATP-binding cassette multidrug transport protein ATRC [Emericella nidulans]		transport	1023, 1026, 1044
2601	78906685 (9419, 9420)	Novel Protein sim. GBank gil5456699 gb AAD43625.1 AF07141 - (AF071410) ATP-binding cassette multidrug transport protein ATRC [Emericella nidulans]		transport	1017, 1022, 1026
2602	79177918 (13781, 13782)	Novel Protein sim. GBank gil5457682 emb CAB4973.1 - (AJ248283) PAB2398 [Pyrococcus abyssi]		transport	1027
2603	79831136 (2549, 2550)	Novel Protein sim. GBank gil5459395 emb CAB50753.1 - (AL096839) putative ABC transporter, ATP-binding subunit [Streptomyces coelicolor]		transport	1006, 1023, 1024, 1034, 1039
2604	79610701 (4821, 4822)	Novel Protein sim. GBank gil5459395 emb CAB50753.1 - (AL096839) putative ABC transporter, ATP-binding subunit [Streptomyces coelicolor]		transport	1034
2605	20686637 (7479, 7480)	Novel Protein sim. GBank gil547880 sp P25737 LYSP_ECOLI - LYSINE-SPECIFIC PERMEASE		transport	1022
2606	20748551 (13629, 13630)	Novel Protein sim. GBank gil547880 sp P25737 LYSP_ECOLI - LYSINE-SPECIFIC PERMEASE		transport	1022
2607	29360922 (18335, 18336)	Novel Protein sim. GBank gil548706 sp P36948 RBSC_BACSU - RIBOSE TRANSPORT SYSTEM PERMEASE PROTEIN RBSC		transport	1009
2608	78361460 (1111, 1112)	Novel Protein sim. GBank gil548909 sp P36559 SDAC_ECOLI - SERINE TRANSPORTER		transport	1008
2609	20634877 (1113, 1114)	Novel Protein sim. GBank gil549650 sp P36171 YK84 - YEAST - PROBABLE ATP-DEPENDENT TRANSPORTER YKR104W		transport	1022

2610	80060202 (18007, 18008)	Novel Protein sim. GBank gil549751 sp P36038 YKU8_YEAST - HYPOTHETICAL 31.2 KD PROTEIN IN STEG- LOS1 INTERGENIC REGION		transport	1008, 1010, 1022, 1024
2611	24132711 (20223, 20224)	Novel Protein sim. GBank gil557058 (U15279) - Webip [Saccharomyces cerevisiae]		transport	1009
2612	66582344 (5087, 5088)	Novel Protein sim. GBank gil5579394 gb AAD45529.1 AF16265 - (AF162655) choline transporter [Streptococcus pneumoniae]		transport	1009
2613	80382826 (15971, 15972)	Novel Protein sim. GBank gil558389 emb CAA86236 - (Z38113) len: 888, CAI: 0.19 incomplete or similar to SNQ2_YEAST P32568 SNQ2 PROTEIN [Saccharomyces cerevisiae]		transport	1001, 1004, 1006, 1010, 1022, 1027, 1029
2614	10870915 (5263, 5264)	Novel Protein sim. GBank gil575393 emb CAA57938 - (X82612) ATM1 [Saccharomyces cerevisiae]		transport	1010
2615	85528233 (5987, 5988)	Novel Protein sim. GBank gil575931 emb CAA57938 - (X82612) ATM1 [Saccharomyces cerevisiae]		transport	1038
2616	79869872 (3425, 3426)	Novel Protein sim. GBank gil585066 sp P37316 DPPB_ECOLI - DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB		transport	1008
2617	25263083 (11913, 11914)	Novel Protein sim. GBank gil585066 sp P37316 DPPB_ECOLI - DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB		transport	1026
2618	37807640 (2153, 2154)	Novel Protein sim. GBank gil585409 sp P22731 LIVF_ECOLI - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT ATP-BINDING PROTEIN LIVF (LIVF PROTEIN F)		transport	1058
2619	10080207 (18391, 18392)	Novel Protein sim. GBank gil585582 sp P38045 NRTC_SYNP7 - NITRATE TRANSPORT ATP-BINDING PROTEIN NRTC		transport	1012

2620	95005558 (1209, 1210)	Novel Protein sim. GBank gi 588638 sp P36318 YB70 YEAST - HYPOTHETICAL 63.0 KD PROTEIN IN PYC2-PDB1 INTERGENIC REGION		transport	1010, 1022, 1034
2621	80237751 (7003, 7004)	Novel Protein sim. GBank gi 588638 sp P38137 FAT2 YEAST - PEROXISOMAL-COENZYME A SYNTHETASE		transport	1001, 1003, 1004, 1006, 1008, 1009, 1017, 1022, 1024, 1026, 1034, 1038, 1039, 1044, 1054
2622	3041958 (16743, 16744)	Novel Protein sim. GBank gi 606034 (U18997) - ORF o472 [Escherichia coli]		transport	1044
2623	27324690 (10845, 10846)	Novel Protein sim. GBank gi 28862 sp P40812 ANSP SALTY - L-ASPARAGINE PERMEASE (L-ASPARAGINE TRANSPORT PROTEIN)		transport	1044
2624	30269312 (16491, 16492)	Novel Protein sim. GBank gi 728904 sp P38929 ATC2 YEAST - CALCIUM-TRANSPORTING ATPASE 2 (VACUOLAR CA2+-ATPASE)		transport	1026
2625	29354390 (21971, 21972)	Novel Protein sim. GBank gi 728904 sp P38929 ATC2 YEAST - CALCIUM-TRANSPORTING ATPASE 2 (VACUOLAR CA2+-ATPASE)		transport	1009
2626	29000310 (21881, 21882)	Novel Protein sim. GBank gi 728906 sp P32660 ATC5 YEAST - PROBABLE CALCIUM-TRANSPORTING ATPASE 5		transport	1044
2627	27824617 (9821, 9822)	Novel Protein sim. GBank gi 729781 sp P38695 HXT5 GLUCOSE TRANSPORTER HXT5		transport	1004
2628	11288675 (3093, 3094)	Novel Protein sim. GBank gi 729784 sp P39924 HXTC YEAST - HEXOSE TRANSPORTER HXT13		transport	1024

2629	17959822 (13449, 13450)	Novel Protein sim. GBank gi 730601 sp P39137 ROCE_BACSU - AMINO- ACID PERMEASE ROCE		transport	1001
2630	79115051 (11953, 11954)	Novel Protein sim. GBank gi 730733 sp P40357 SEC9_YEAST - PROTEIN TRANSPORT PROTEIN SEC9		transport	1003
2631	78902702 (22495, 22496)	Novel Protein sim. GBank gi 730755 sp P38778 SMF2_YEAST - TRANSPORTER PROTEIN SMF2		transport	1003, 1008, 1009, 1017, 1026
2632	65899490 (19261, 19262)	Novel Protein sim. GBank gi 731052 sp P10907 UGPC_ECOLI - SN- GLYCEROL-3-PHOSPHATE TRANSPORT ATP- BINDING PROTEIN UGPC		transport	1044
2633	32567151 (4071, 4072)	Novel Protein sim. GBank gi 731430 sp P39973 YEI4_YEAST - HYPOTHETICAL 13.1 KD PROTEIN IN HXT8 5REGION		transport	1034
2634	78937538 (14283, 14284)	Novel Protein sim. GBank gi 731625 sp P38722 YHE9_YEAST - HYPOTHETICAL 31.5 KD PROTEIN IN CBP2 5REGION		transport	1008, 1009, 1017, 1022, 1026, 1044
2635	57293328 (11307, 11308)	Novel Protein sim. GBank gi 731645 sp P38767 YHJ2_YEAST - HYPOTHETICAL 64.2 KD PROTEIN IN SLT2- PUT2 INTERGENIC REGION		transport	1004, 1006, 1010, 1017, 1022, 1024, 1034
2636	5821661 (22175, 22176)	Novel Protein sim. GBank gi 731645 sp P38767 YHJ2_YEAST - HYPOTHETICAL 64.2 KD PROTEIN IN SLT2- PUT2 INTERGENIC REGION		transport	1058
2637	36827480 (7277, 7278)	Novel Protein sim. GBank gi 731893 sp P40445 YIQ6_YEAST - PUTATIVE TRANSPORTER YIL166C		transport	1054

2638	80499939 (9121, 9122)	Novel Protein sim. GBank gi 732208 sp P27514 YN86_YEAST - HYPOTHETICAL 99.5 KD PROTEIN IN URK1- SMM1 INTERGENIC REGION	transport	1001, 1003, 1004, 1006, 1009, 1010, 1012, 1022, 1023, 1024, 1025, 1026, 1027, 1034, 1044
2639	87898857 (13311, 13312)	Novel Protein sim. GBank gi 732230 sp P40979 YOR1_CALSR - HYPOTHETICAL PROTEIN IN XYLR 5'REGION (ORF1)	transport	1015, 1026
2640	65691891 (8395, 8396)	Novel Protein sim. GBank gi 732281 sp P39312 CYCA_ECOLI -D- SERINE/D-ALANINE/GLYCINE TRANSPORTER	transport	1009
2641	987561 (11677, 11678)	Novel Protein sim. GBank gi 732295 sp P39326 YTFR_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP- BINDING PROTEIN IN PPA-FBP INTERGENIC REGION	transport	1044
2642	10047630 (8029, 8030)	Novel Protein sim. GBank gi 732297 sp P39328 YTFT_ECOLI - HYPOTHETICAL ABC TRANSPORTER PERMEASE PROTEIN YTFT	transport	1039
2643	79456297 (2727, 2728)	Novel Protein sim. GBank gi 81286 pir S22697 - extensin - Volvox carteri (fragment)	transport	1006, 1026
2644	80079967 (7371, 7372)	Novel Protein sim. GBank gi 868021 M14641 - glucuronidase permease [Escherichia coli]	transport	1030, 1034
2645	33204860 (2113, 2114)	Novel Protein sim. GBank gi 97480 pir S19739 - integral membrane protein - Rhodobacter capsulatus	transport	1026
2646	78804698 (2395, 2396)		transport	1030, 1038
2647	80026653 (11979, 11980)		transport	1006, 1034

2648	30656290 (1521, 1522)	Novel Protein sim. GBank gil135411sp P14641 TBA2_MAIZE - TUBULIN ALPHA-2 CHAIN		tubulin	1026
2649	29352720 (15311, 15312)	Novel Protein sim. GBank gil135411sp P14641 TBA2_MAIZE - TUBULIN ALPHA-2 CHAIN		tubulin	1022, 1023, 1044
2650	85822318 (14369, 14370)	Novel Protein sim. GBank gil2670711sp P30436 TBA_ONCKE - TUBULIN ALPHA CHAIN		tubulin	1019
2651	86695830 (14947, 14948)	Novel Protein sim. GBank gil2670711sp P29514 TBB6_ARATH - TUBULIN BETA-6 CHAIN		tubulin	1003, 1008, 1009, 1016, 1017, 1022, 1026, 1031, 1044
2652	78674822 (17031, 17032)	Novel Protein sim. GBank gil401161sp Q02245 TBA5_MAIZE - TUBULIN ALPHA-5 CHAIN		tubulin	1009
2653	11371791 (6445, 6446)	Novel Protein sim. GBank gil401439sp P31386 ATS1_YEAST - ATS1 PROTEIN (ALPHA-TUBULIN SUPPRESSOR 1)		tubulin	1022, 1024
2654	78795118 (22763, 22764)	Novel Protein sim. GBank gil401439sp P31386 ATS1_YEAST - ATS1 PROTEIN (ALPHA-TUBULIN SUPPRESSOR 1)		tubulin	1026
2655	19847809 (8221, 8222)	Novel Protein sim. GBank gil417049sp Q03555 GEPH_RAT - GEPHYRIN (PUTATIVE GLYCINE RECEPTOR-TUBULIN LINKER PROTEIN)		tubulin	1004, 1022
2656	78895987 (4117, 4118)	Novel Protein sim. GBank gil918581pri D23437 - tubulin beta-4 chain - mouse		tubulin	1008
2657	78138540 (14927, 14928)	Novel Protein sim. GBank gil4507855refNP_003472.1 pUSP5 - Ubiquitin isopeptidase T	Contains protein domain (PF0442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	1000, 1029, 1034

2658	79181325 (14647, 14648)	Novel Protein sim. GBank gi 5884991 sp P38237 UBPE_YEAST - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 14 (UBIQUITIN THIOLESTERASE 14) (UBIQUITIN- SPECIFIC PROCESSING PROTEASE 14) (DEUBIQUITINATING ENZYME 14)	Contains protein domain (PF00442) - Ubiquitin carboxyl-terminal hydrolases family 2	ubiquitin	1008, 1009, 1026, 1031, 1039, 1044
2659	87113899 (5903, 5904)	Novel Protein sim. GBank gi 11748621 sp Q09738 UBPA_SCHPO - PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C13A11.04C (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	1048
2660	78763202 (11959, 11960)	Novel Protein sim. GBank gi 4469352 gb AAAD21222 - (AF069502) ubiquitin specific protease UBP43 [Mus musculus]	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	1009, 1030
2661	80217694 (11881, 11882)	Novel Protein sim. GBank gi 731042 sp P39944 UBP5_YEAST - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 5 (UBIQUITIN THIOLESTERASE 5) (UBIQUITIN- SPECIFIC PROCESSING PROTEASE 5) (DEUBIQUITINATING ENZYME 5)	Contains protein domain (PF00443) - Ubiquitin carboxyl-terminal hydrolase family 2	ubiquitin	1004, 1006, 1022, 1024, 1025
2662	57532338 (1953, 1954)	Novel Protein sim. GBank gi 4584342 gb AAAD25137.1 AC00712 - (AC007127) putative ubiquitin protein [Arabidopsis thaliana]	Contains protein domain (PF00627) - UBA domain	ubiquitin	1058
2663	71841342 (10015, 10016)	Novel Protein sim. GBank gi 136684 sp P19812 UBR1_YEAST - N-END- RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-RECOGNIN)		ubiquitin	1008, 1026

2664	79818307 (17375) 17376)	Novel Protein sim. GBank gi 136684 sp P19812 UBR1_YEAST - N-END- RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-RECOGNIN)		ubiquitin	1006, 1016
2665	78731537 (20935, 20936)	Novel Protein sim. GBank gi 1717871 sp P50101 UBPF_YEAST - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 15 (UBIQUITIN THIOLESTERASE 15) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 15) (DEUBIQUITINATING ENZYME 15)		ubiquitin	1008
2666	66694767 (3579, 3580)	Novel Protein sim. GBank gi 2131505 pir S69625 - hypothetical protein YDR457w - yeast (Saccharomyces cerevisiae)		ubiquitin	1009
2667	80496584 (19751, 19752)	Novel Protein sim. GBank gi 2132663 pir S64851 - probable membrane protein YLR024c - yeast (Saccharomyces cerevisiae)		ubiquitin	1003, 1004, 1006, 1008, 1009, 1010, 1012, 1016, 1017, 1022, 1023, 1024, 1025, 1027, 1031, 1034, 1038, 1044
2668	23332035 (18771, 18772)	Novel Protein sim. GBank gi 2190550 (AC001229) - ESTs gi T45673 gb N37512 come from this gene. [Arabidopsis thaliana]		ubiquitin	1053
2669	80504149 (4689, 4690)	Novel Protein sim. GBank gi 2842691 sp Q92353 UBPC_SCHPO - PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08 (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)		ubiquitin	1012, 1022, 1023, 1038

2670	20435987 (14725, 14726)	Novel Protein sim. GBank gil3184080 embj CAA19336I - (AL023781 hypothetical protein [Schizosaccharomyces pombe])		ubiquitin	1024
2671	34083152 (9893, 9894)	Novel Protein sim. GBank gil3386552 (AF079565) - ubiquitin-specific protease UBP41 [Mus musculus]		ubiquitin	1000
2672	20730990 (10917, 10918)	Novel Protein sim. GBank gil401237 sp P31251 UBA2_WHEAT - UBIQUITIN ACTIVATING ENZYME E1_2		ubiquitin	1022
2673	80079058 (18923, 18924)	Novel Protein sim. GBank gil4580007 dbj BAA75198.1I - (D87259) poly(A)+ RNA transport protein Ptr3p [Schizosaccharomyces pombe]		ubiquitin	1034
2674	26249117 (14729, 14730)	Novel Protein sim. GBank gil568303 sp P38290 UBS1_YEAST - UBIQUITIN-CONJUGATING ENZYME SUPPRESSOR 1		ubiquitin	1003
2675	29238694 (12989, 12990)	Novel Protein sim. GBank gil731042 sp P3944 UBP5_YEAST - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 5 (UBIQUITIN THIOLESTERASE 5) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 5) (DEUBIQUITINATING ENZYME 5)		ubiquitin	1017
2676	78373468 (23, 24)	Novel Protein sim. GBank gil731045 sp P39538 UBPC_YEAST - UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 12 (UBIQUITIN THIOLESTERASE 12) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 12) (DEUBIQUITINATING ENZYME 12)		ubiquitin	1031
2677	80090632 (21941, 21942)	Novel Protein sim. GBank gil4504091 ref NP_001496.1 pGPR3 - G protein-coupled receptor 30	Contains protein domain (PF00001) - 7 transmembrane receptor (rhodopsin family)	UNCLASSIFIED	1022, 1024, 1040, 1049

2678	35103778 (21187, 21188)	Novel Protein sim. GBank g 2244797 emb CAB10220.1 - (Z97336) [hypothetical protein [Arabidopsis thaliana]]	Contains protein domain (PF00010) - Helix-loop-helix DNA-binding domain	UNCLASSIFIED	1034
2679	13883867 (19495, 19496)	Novel Protein sim. GBank g 2313186 gb AAD07173.1 - (AE000532) methyl-accepting chemotaxis protein (tppB) [Helicobacter pylori 26835]	Contains protein domain (PF00015) - Methyl-accepting chemotaxis protein (MCP) signaling domain	UNCLASSIFIED	1053
2680	56066458 (18383, 18384)	Novel Protein sim. GBank g 483202 pir D45835 - Ly6 homolog Rk3 precursor - rat	Contains protein domain (PF00021) - u-FAR/Ly-6 domain	UNCLASSIFIED	1010
2681	78728633 (6693, 6694)	Novel Protein sim. GBank g 114124 sp P19146 ARF2_YEAST - ADP-RIBOSYLATION FACTOR 2	Contains protein domain (PF00025) - ADP-ribosylation factor family	UNCLASSIFIED	1022
2682	66685502 (13921, 13922)	Novel Protein sim. GBank g 2072425 (U83115) - non-lens beta crystallin like protein [Homo sapiens]	Contains protein domain (PF00030) - Beta/Gamma crystallin	UNCLASSIFIED	1010
2683	56726485 (3209, 3210)	Novel Protein sim. GBank g 1572819 (U70855) - similar to the RAS gene family [Caenorhabditis elegans]	Contains protein domain (PF00036) - EF hand	UNCLASSIFIED	1034
2684	10156992 (11089, 11090)	Novel Protein sim. GBank g 2811081 sp P00208 FER_CHRVI - FERREDOXIN	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.	UNCLASSIFIED	1003
2685	20700110 (4297, 4298)	Novel Protein sim. GBank g 1076211 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii	Contains protein domain (PF00040) - Fibronectin type II domain	UNCLASSIFIED	1022
2686	85533136 (8059, 8060)	Novel Protein sim. GBank g 731924 sp P40582 YIV8_YEAST - HYPOTHETICAL 26.8 KD PROTEIN IN HYR1 3REGION	Contains protein domain (PF00043) - Glutathione S-transferases.	UNCLASSIFIED	1049
2687	86674019 (16335, 16336)	Novel Protein sim. GBank g 413813 emb CAA09791 - (A_011828) NDX1 homeobox protein [Lotus japonicus]	Contains protein domain (PF00046) - Homeobox domain	UNCLASSIFIED	1054

2688	55777970 (22403, 22404)	Novel Protein sim. GBank gil4099555 (U88678) - immunoglobulin superfamily member [Drosophila melanogaster]	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	1021
2689	99419769 (22501, 22502)	Novel Protein sim. GBank gil4759066 ref NP_004579.1 pSCN2 - sodium channel, voltage-gated, type II, beta polypeptide	Contains protein domain (PF00047) - Immunoglobulin domain	UNCLASSIFIED	1017, 1040
2690	20443181 (16665, 16666)		Contains protein domain (PF00056) - lactate/malate dehydrogenase	UNCLASSIFIED	1024
2691	65876060 (1201, 1202)	Novel Protein sim. GBank gil127533 sp P02761 MUP_RAT - MAJOR URINARY PROTEIN PRECURSOR (MUP) (ALPHA-2U-GLOBULIN) (15.5 KD FATTY ACID BINDING PROTEIN) (15.5 KD FABP)	Contains protein domain (PF00061) - lipocalin	UNCLASSIFIED	1022, 1026, 1034, 1044
2692	79915075 (14267, 14268)		Contains protein domain (PF00067) - Cytochrome P450	UNCLASSIFIED	1017
2693	79254687 (9107, 9108)	Novel Protein sim. GBank gil4506537 ref NP_003812.1 pRIP2 - UNKNOWN	Contains protein domain (PF00069) - Eukaryotic Protein Kinase domain	UNCLASSIFIED	1054
2694	27844242 (8627, 8628)	Novel Protein sim. GBank gil861110 emb CAA61012 - (X87672) ALK-1 (Saccharomyces cerevisiae)	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	UNCLASSIFIED	1022, 1024, 1027
2695	11759517 (14753, 14754)	Novel Protein sim. GBank gil3282794 (AF044495) - CheB homolog [Agrobacterium tumefaciens]	Contains protein domain (PF00072) - Response regulator receiver domain	UNCLASSIFIED	1024
2696	80035435 (16091, 16092)	Novel Protein sim. GBank gil3294237 emb CAA19850 - (AL031031) two-component regulator [Streptomyces coelicolor]	Contains protein domain (PF00072) - Response regulator receiver domain	UNCLASSIFIED	1053
2697	38406286 (8117, 8118)	Novel Protein sim. GBank gil2500731 sp Q61247 A2AP_MOUSE - ALPHA-2-ANTIPLASMIN PRECURSOR (ALPHA-2-PLASMIN INHIBITOR) (ALPHA-2-P) (ALPHA-2-AP)	Contains protein domain (PF00079) - Serpins (serine protease inhibitors)	UNCLASSIFIED	1008, 1009

2698	80502156 (6103, 6104)		Contains protein domain (PF00096) - Zinc finger, C2H2 type	UNCLASSIFIED	1012
2699	29141248 (15535, 15536)	Novel Protein sim. GBank gi 1078084 pir S53414 probable membrane protein YLR427w - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	1017
2700	21434586 (10341, 10342)	Novel Protein sim. GBank gi 2131354 pir S61037 hypothetical protein YDL175C - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00098) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	1017, 1034
2701	80046161 (12313, 12314)	Novel Protein sim. GBank gi 2244916 emb CAB10338_11 - (Z97339) hypothetical protein [Arabidopsis thaliana]	Contains protein domain (PF00098) - Zinc finger, C3HC class	UNCLASSIFIED	1010
2702	47656346 (9085, 9086)	Novel Protein sim. GBank gi 3435302 (AF082873) - glutamine synthetase [Herbaspirillum seropedicae]	Contains protein domain (PF00120) - Glutamine synthetase	UNCLASSIFIED	1029
2703	20228021 (1135, 1136)	Novel Protein sim. GBank gi 1050749 emb CAA61941 - (X89816) GSTR protein [Rhizobium leguminosarum]	Contains protein domain (PF00126) - Bacterial regulatory helix-turn-helix protein, lysR family	UNCLASSIFIED	1022
2704	79842462 (17959, 17960)		Contains protein domain (PF00127) - Copper binding proteins, plastocyanin/azurin family	UNCLASSIFIED	1044
2705	13883320 (13439, 13440)	Novel Protein sim. GBank gi 2313283 gb AAD07263_1 - (AE0000539) UDP-3-O-(3-hydroxymyristoyl) glucosamine N-acetyltransferase (ipxD) [Helicobacter pylori 26695]	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	UNCLASSIFIED	1053
2706	11614001 (7855, 7856)	Novel Protein sim. GBank gi 2642159 (AC003000) - putative mannose-1'-phosphate guanylyltransferase [Arabidopsis thaliana]	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	UNCLASSIFIED	1016
2707	78385390 (11857, 11858)	Novel Protein sim. GBank gi 7294661 sp P40882 FBP_PSEAE - FERRIPYOCHELIN BINDING PROTEIN	Contains protein domain (PF00132) - Bacterial transferase hexapeptide (four repeats)	UNCLASSIFIED	1026

2708	79462845 (1199, 1200)	Novel Protein sim. GBank gi 3261828 emb CAB10925 - (298260) mpr [Mycobacterium tuberculosis]	Contains protein domain (PF00142) - 4Fe-4S iron sulfur cluster binding proteins. NifH/ftrxC family	UNCLASSIFIED	1024, 1026, 1027, 1031
2709	80570822 (20817, 20818)	Novel Protein sim. GBank gi 1272761 sp P16036 MPCP_RAT - MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN PRECURSOR	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	1058
2710	78169281 (4331, 4332)	Novel Protein sim. GBank gi 2052355 U94592 - uncoupling protein homolog [Homo sapiens]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	1053
2711	35893437 (20833, 20834)	Novel Protein sim. GBank gi 2408051 emb CAB16300 - (299168) putative RNA splicing protein [Schizosaccharomyces pombe]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	1053
2712	30635867 (9241, 9242)	Novel Protein sim. GBank gi 3879938 emb CAA92487 - (268220) Similarity to Human ADP/ATP carrier protein (SW:ADT1_HUMAN); cDNA EST EMBL:D71893 comes from this gene; cDNA EST EMBL:D74630 comes from this gene; cDNA EST yk394f1.3 comes from this gene; cDNA EST yk394f1.5 comes from ...	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	1022
2713	65864785 (17679, 17680)	Novel Protein sim. GBank gi 485131 U00052 - similar to ADP, ATP carrier proteins [Caenorhabditis elegans]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	UNCLASSIFIED	1016, 1017
2714	78462780 (111349, 111350)	Novel Protein sim. GBank gi 1703350 sp P52561 APT_STRCO - ADENINE PHOSPHORIBOSYLTRANSFERASE	Contains protein domain (PF00156) - Phosphoribosyl transferase domain	UNCLASSIFIED	1026
2715	80257955 (22429, 22430)	Novel Protein sim. GBank gi 2632508 emb CAB12016 - (299105) similar to transcriptional regulator (AraC/XylS family) [Bacillus subtilis]	Contains protein domain (PF00165) - Bacterial regulatory helix-turn-helix proteins, araC family	UNCLASSIFIED	1006

2716	57517767 (5049, 5050)	Novel Protein sim. GBank gil1170163 sp P41546 HAC1_YEAST - HAC1 PROTEIN	Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	1001, 1003, 1004, 1006, 1008, 1010, 1012, 1016, 1017, 1022, 1024, 1026, 1027, 1029, 1031, 1034
2717	80222170 (8635, 8636)		Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	1001
2718	15025261 (12765, 12766)	Novel Protein sim. GBank gil320712 p rl S30839 - UTR2 protein - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00187) - Chitin recognition protein	UNCLASSIFIED	1054
2719	39516164 (12525, 12526)	Novel Protein sim. GBank gil345658 sp P48284 CAH4_RAT - CARBONIC ANHYDRASE IV PRECURSOR (CARBONATE DEHYDRATASE IV)	Contains protein domain (PF00194) - Eukaryotic-type carbonic anhydrase	UNCLASSIFIED	1006
2720	30455609 (20323, 20324)		Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal-phosphate	UNCLASSIFIED	1008
2721	35067177 (14607, 14608)	Novel Protein sim. GBank gil117698 sp P42415 OLD_BACSU - IOLD PROTEIN	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	UNCLASSIFIED	1054
2722	80499903 (16571, 16572)	Novel Protein sim. GBank gil117698 sp P42415 OLD_BACSU - IOLD PROTEIN	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	UNCLASSIFIED	1008, 1012
2723	79834671 (9265, 9266)	Novel Protein sim. GBank gil2622556 (AE000905) - acetolactate synthase, large subunit [Methanobacterium thermoautotrophicum]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	UNCLASSIFIED	1039
2724	78461385 (2657, 2658)	Novel Protein sim. GBank gil1311484 dbj BAA12702 - (D84646) fumarase [Thermus aquaticus]	Contains protein domain (PF00206) - Lyase	UNCLASSIFIED	1003

2725 129962)	20728701 (12961, gii 31635 sp P12047 PUR8_BACSU - ADENYLOSUCCINATE LYASE (ADENYLOSUCCINASE) (ASL)	Novel Protein sim. GBank gii 289673 emb CAA1724.4.1 - (AL021897) fum [Mycobacterium tuberculosis]	Contains protein domain (PF00206) - Lyase	UNCLASSIFIED	1004, 1006
2726 14960)	47658232 (14959, 14960)	Novel Protein sim. GBank gii 134036 sp P23887 FTN_ECOLI - FERRITIN	Contains protein domain (PF00206) - Lyase	UNCLASSIFIED	1029
2727 9074)	80250842 (9073, 9074)	Novel Protein sim. GBank gii 134036 sp P23887 FTN_ECOLI - FERRITIN	Contains protein domain (PF00210) - Ferritins	UNCLASSIFIED	1022, 1029
2728 15494)	79186400 (15493, 15494)	Novel Protein sim. GBank gii 341341 emb CAA20279 - (AL031232) [hypotheical protein SC10H5.07 [Streptomyces coelicolor]]	Contains protein domain (PF00220) - Neurohypophysis hormones, N-terminal Domain	UNCLASSIFIED	1027
2729 7742)	68983289 (7741, 7742)		Contains protein domain (PF00220) - Neurohypophysis hormones, N-terminal Domain	UNCLASSIFIED	1021, 1046
2730 13092)	80413902 (13091, 13092)		Contains protein domain (PF00224) - Pyruvate kinase	UNCLASSIFIED	1006, 1024, 1025, 1027, 1034, 1039
2731 15088)	88095873 (15087, 15088)	Novel Protein sim. GBank gii 3367516 (AC004392) - Similar to beta-glucosidase BGQ60 precursor gb L418669 from Hordeum vulgare. [Arabidopsis thaliana]	Contains protein domain (PF00232) - Glycosyl hydrolase family 1	UNCLASSIFIED	1003
2732 9202)	79821878 (9201, 9202)	Novel Protein sim. GBank gii 1100235 (L48985) - resolvase [Pseudomonas syringae]	Contains protein domain (PF00239) - Site-specific recombinases	UNCLASSIFIED	1034
2733 16062)	79456375 (16061, 16062)	Novel Protein sim. GBank gii 154917 (M95402) - resolvase [Transposon Tn5393]	Contains protein domain (PF00239) - Site-specific recombinases	UNCLASSIFIED	1031
2734 3450)	80075605 (3449, 3450)	Novel Protein sim. GBank gii 178117 emb CAB06106 - (Z83858) [hypothetical protein Rv2843 [Mycobacterium tuberculosis]]	Contains protein domain (PF00239) - Site-specific recombinases	UNCLASSIFIED	1006, 1034

2735	79614907 (19739, 19740)	Novel Protein sim. GBank gil12506785 sp P03014 DNIV_ECOLI - DNA-INVERTASE PIN	Contains protein domain (PF00239) - Site-specific recombinases	UNCLASSIFIED	1003, 1004, 1006
2736	65484397 (595, 596)	Novel Protein sim. GBank gil1723683 sp P5325 YG2H YEAST - HYPOTHETICAL 37.1 KD PROTEIN IN PAC10-TOM20 INTERGENIC REGION	Contains protein domain (PF00241) - Cofilin/tropomyosin-type actin-binding proteins	UNCLASSIFIED	1023
2737	27842921 (6857, 6858)	Novel Protein sim. GBank gil244239 bbs 78626 - (S78624) YCR592 [Saccharomyces cerevisiae=yeast, Peptide, 1226 aa] [Saccharomyces cerevisiae]	Contains protein domain (PF00249) - Myb-like DNA-binding domain	UNCLASSIFIED	1004, 1010
2738	86475875 (16741, 16742)		Contains protein domain (PF00254) - FKBp-type peptidyl-prolyl cis-trans isomerases	UNCLASSIFIED	1029
2739	98201111 (2959, 2960)	Novel Protein sim. GBank gil10773 2 pir S50953 - hypothetical protein YLL066c - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	1004, 1010, 1022, 1024, 1034, 1044, 1058
2740	95290543 (10641, 10642)	Novel Protein sim. GBank gil2506493 sp P38036 YGCB_ECOLI - HYPOTHETICAL 100.5 KD PROTEIN IN IAP-CYSH INTERGENIC REGION	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	1004, 1006, 1010, 1022, 1023, 1024, 1031, 1034, 1039, 1049, 1054
2741	988205 (8891, 8892)	Novel Protein sim. GBank gil113580 sp P02770 ALBU_RAT - SERUM ALBUMIN PRECURSOR	Contains protein domain (PF00273) - Serum albumin family	UNCLASSIFIED	1008
2742	80505192 (18093, 18094)	Novel Protein sim. GBank gil584743 sp Q07158 ALF_STACA - FRUCTOSE-BISPHOSPHATE ALDOLASE CLASS I (FBP ALDOLASE)	Contains protein domain (PF00274) - Fructose-bisphosphate aldolase class-I	UNCLASSIFIED	1003, 1004, 1006, 1009, 1012, 1017, 1022, 1024, 1039
2743	39523995 (18935, 18936)	Novel Protein sim. GBank gil3123042 sp Q06208 Y1ER_MYCTU - HYPOTHETICAL 31.4 KD PROTEIN CY1A10.27	Contains protein domain (PF00290) - Tryptophan synthase alpha chain	UNCLASSIFIED	1004

2744	80079069 (20221, 20222),	Novel Protein sim. GBank g 703022 sp P53368 8ODP_MOUSE - 7,8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (8-OXO-DGTPASE)	Contains protein domain (PF00293) - Bacterial mutT protein	UNCLASSIFIED	1034
2745	86475395 (12299, 12300)	Novel Protein sim. GBank g 2497998 sp P95781 MUTX_STRMU - MUTATOR MUTT PROTEIN (7,8-DIHYDRO-8-OXOGUANINE-TRIPHOSPHATASE) (8-OXO-DGTPASE) (DGTP PYROPHOSPHOHYDROLASE)	Contains protein domain (PF00293) - Bacterial mutT protein	UNCLASSIFIED	1008, 1048
2746	27218037 (18689, 18690)	Novel Protein sim. GBank g 282760 emb CAA16664 - (AL021646) hypothetical protein Rv3199c [Mycobacterium tuberculosis]	Contains protein domain (PF00293) - Bacterial mutT protein	UNCLASSIFIED	1009
2747	11772628 (515, 516)	Novel Protein sim. GBank g 132127 sp P05054 RBSK_ECOLI - RIBOKINASE	Contains protein domain (PF00294) - pfkB family carbohydrate kinase	UNCLASSIFIED	1022
2748	78385176 (17953, 17954)		Contains protein domain (PF00300) - Phosphoglycerate mutase family	UNCLASSIFIED	1026
2749	11751363 (17081, 17082)	Novel Protein sim. GBank g 2494413 sp Q59758 DNAA_RICPR - CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA	Contains protein domain (PF00308) - Bacterial dnaA protein	UNCLASSIFIED	1022
2750	47656860 (1681, 1682)	Novel Protein sim. GBank g 3800785 AF071024 - DNA replication Initiator protein [Streptomyces chrysomallus]	Contains protein domain (PF00308) - Bacterial dnaA protein	UNCLASSIFIED	1029
2751	88097357 (19651, 19652)	Novel Protein sim. GBank g 39533448 U19185 - DnaA [Mycobacterium avium]	Contains protein domain (PF00308) - Bacterial dnaA protein	UNCLASSIFIED	1018
2752	47658191 (355, 356)	Novel Protein sim. GBank g 4097540 U64826 - glutamine phosphoribosylpyrophosphate amidotransferase [Streptomyces lividans]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	UNCLASSIFIED	1029
2753	65642511 (1763, 1764)	Novel Protein sim. GBank g 4097540 U64826 - glutamine phosphoribosylpyrophosphate amidotransferase [Streptomyces lividans]	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	UNCLASSIFIED	1016

2754	86473966 (17739) 17740)	Novel Protein sim. GBank gil542200 pir S41890 - hypothetical protein - garden asparagus	Contains protein domain (PF00310) - Glutamine amidotransferases class-II	UNCLASSIFIED	1054
2755	80218800 (7959, 7960)	Novel Protein sim. GBank gil4981768 gb AAD36290.1 AE00177 - (AE001778) NADH dehydrogenase, 30 kDa subunit, putative [Thermotoga maritima]	Contains protein domain (PF00329) - Respiratory-chain NADH dehydrogenase, 30 Kd subunit	UNCLASSIFIED	1008, 1010, 1022, 1023, 1026, 1030, 1031, 1038, 1053
2756	25176960 (4915, 4916)	Novel Protein sim. GBank gil1351856 sp P49608 ACOC_CUCMA - ACONITATE HYDRATASE, CYTOPLASMIC (CITRATE HYDRO-LYASE) (ACONITASE)	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	UNCLASSIFIED	1003
2757	20297928 (17879, 17880)	Novel Protein sim. GBank gil2791409 emb CAA160031 - (AL021184) acn [Mycobacterium tuberculosis]	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	UNCLASSIFIED	1034
2758	65496920 (10007, 10008)	Novel Protein sim. GBank gil2233084 emb CAA746021 - (Y14206) orfsyn [Streptomyces coelicolor]	Contains protein domain (PF00334) - Nucleoside diphosphate Kinases	UNCLASSIFIED	1023
2759	80062845 (15999, 16000)	Novel Protein sim. GBank gil729490 sp P31697 FIMC_ECOLI - CHAPERONE PROTEIN FIMC PRECURSOR	Contains protein domain (PF00345) - Gram-negative pili assembly chaperone	UNCLASSIFIED	1010
2760	37816247 (11563, 11564)	Novel Protein sim. GBank gil7317284 sp P39722 YAE8_YEAST - HYPOTHETICAL 75.2 KD PROTEIN IN ACS1- GCV3 INTERGENIC REGION	Contains protein domain (PF00350) - Dynamin family	UNCLASSIFIED	1008, 1058
2761	39710223 (589, 590)	Novel Protein sim. GBank gil2905647 (AF045245) - D-arabinitol kinase [Klebsiella pneumoniae]	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	UNCLASSIFIED	1008
2762	79814894 (19441, 19442),	Novel Protein sim. GBank gil2126549 pir 399692 - [NiFe] uptake hydrogenase large subunit - Anabaena sp.	Contains protein domain (PF00374) - Nickel-dependent hydrogenases	UNCLASSIFIED	1017, 1038
2763	20466360 (10667, 10668)	Novel Protein sim. GBank gil2578382 emb CAA154671 - (AL008609) hypothetical protein MLCB1788.35c [Mycobacterium leprae]	Contains protein domain (PF00376) - Bacterial regulatory proteins, merR family	UNCLASSIFIED	1010

2764	79568388 (11945, 11946)	Novel Protein sim. GBank gii 456867 emb CAB3710 - (A J224978) ORF 319 [Salmonella typhimurium]	Contains protein domain (PF00376) - Bacterial regulatory proteins, merR family	UNCLASSIFIED	1058
2765	11759551 (22185, 22196)	Novel Protein sim. GBank gii 296017 emb CAA18074.1 - (AL022121) [hypothetical protein Rv3752c [Mycobacterium tuberculosis]]	Contains protein domain (PF00383) - Cytidine and deoxycytidylate deaminase zinc-binding region	UNCLASSIFIED	1024
2766	27965981 (1265, 1266)	Novel Protein sim. GBank gii 4757970 ref NP_004816.1 pCDY1 - chromodomain protein, Y chromosome-like	Contains protein domain (PF00385) - 'chromo' (CHRRomatin Organization Modifier) domain	UNCLASSIFIED	1034
2767	79606589 (1275, 1276)	Novel Protein sim. GBank gii 346891 sp P45597 PTF1_XANCP - MULTIPHOSPHORYL TRANSFER PROTEIN (MTP) CONTAINS: PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I); PHOSPHOCARRIER PROTEIN HPR (PROTEIN H); PTS SYSTEM, FRUCTOSE-SPECIFIC II A COMPONENT ...	Contains protein domain (PF00391) - PEP-utilizing enzymes	UNCLASSIFIED	1003
2768	20433599 (1347, 1348)	Novel Protein sim. GBank gii 48515 sp P32670 PT1A_ECOLI - PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE PTS A (PHOSPHOTRANSFERASE SYSTEM, ENZYME I) (ENZYME I-ANI)	Contains protein domain (PF00391) - PEP-utilizing enzymes	UNCLASSIFIED	1010
2769	11245783 (20751, 20752)	Novel Protein sim. GBank gii 1076211 pir S50755 - hypothetical protein VSP-3 - Chlamydomonas reinhardtii	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1034
2770	32338578 (10191, 10192)	Novel Protein sim. GBank gii 4007683 emb CAA22369 - (AL034443) putative transcriptional regulator [Streptomyces coelicolor]	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1029

				UNCLASSIFIED	1031
2771	13042669 (18907, 18908)	Novel Protein sim. GBank g 4106587 emb CAA21342 - (AL031866) ORF19, len: 473 aa, similar to tyrosine aminotransferase, highly similar to E. coli D90784_2, Fasta scores opt:1842, E(): 0 [Yersinia pestis]	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1003, 1008, 1013, 1016, 1025, 1029, 1038, 1053
2772	86471594 (4995, 4996)	Novel Protein sim. GBank g 419140 gb AAD09767 - (AF096293) sporulation transcription factor [Streptomyces aureofaciens]	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1003, 1008, 1013, 1016, 1025, 1029, 1038, 1053
2773	11592156 (14161, 14162)	Novel Protein sim. GBank g 732097 sp P39389 YJIR_ECOLI - HYPOTHETICAL 53.0 KD PROTEIN IN IADA-MCRD INTERGENIC REGION (F470)	Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1034, 1039
2774	79757687 (7877, 7878)		Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1022, 1044
2775	79610951 (14525, 14526)		Contains protein domain (PF00392) - Bacterial regulatory proteins, gntR family	UNCLASSIFIED	1017
2776	80073373 (10365, 10366)	Novel Protein sim. GBank g 2494905 sp Q12024 YTM1_YEAST - MICROTUBULE-ASSOCIATED PROTEIN YTM1	Contains protein domain (PF00400) - WD domain, G-beta repeat	UNCLASSIFIED	1008, 1009, 1034
2777	57108751 (21855, 21856)	Novel Protein sim. GBank g 134913 sp P023391SS_ECOLI - SINGLE-STRAND BINDING PROTEIN (SSB) (HELIX-DESTABILIZING PROTEIN)	Contains protein domain (PF00436) - Single-strand binding protein family	UNCLASSIFIED	1008, 1029, 1034
2778	79556682 (5941, 5942)	Novel Protein sim. GBank g 2983549 (AE000721) - transcriptional regulator [TetR/AcrR family] [Aequifex aeolicus]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	UNCLASSIFIED	1044
2779	79610030 (5451, 5452)		Contains protein domain (PF00445) - Ribonuclease T2 family	UNCLASSIFIED	1003

2780	79843119 (19845, 19946)		Contains protein domain (PF00446) - Gonadotropin- releasing hormones	UNCLASSIFIED	1008, 1024
2781	79579270 (18019, 18020)	Novel Protein sim. GBank gi 3294423 emb CAAI9856 - (AL031031) hypothetical protein SC7C7.10 [Streptomyces coelicolor]	Contains protein domain (PF00455) - Bacterial regulatory proteins, deoR family	UNCLASSIFIED	1001, 1006, 1031
2782	79868677 (15497, 15498)	Novel Protein sim. GBank gi 2553125 (AC02387) - putative transketolase precursor [Arabidopsis thaliana]	Contains protein domain (PF00456) - Transketolase	UNCLASSIFIED	1038
2783	80504814 (8591, 6592)		Contains protein domain (PF00462) - Glutaredoxin	UNCLASSIFIED	1012
2784	82062056 (6909, 6910)	Novel Protein sim. GBank gi 113033 sp P28240 ACEA_YEAST - ISOCITRATE LYASE (ISOCITRASE) (ISOCITRATASE) (ICL)	Contains protein domain (PF00463) - Isocitrate lyase	UNCLASSIFIED	1004, 1006, 1008, 1010, 1016, 1017, 1022, 1024, 1025, 1034, 1044
2785	82457415 (12385, 12386)	Novel Protein sim. GBank gi 113033 sp P28240 ACEA_YEAST - ISOCITRATE LYASE (ISOCITRASE) (ISOCITRATASE) (ICL)	Contains protein domain (PF00463) - Isocitrate lyase	UNCLASSIFIED	1001, 1003, 1004, 1006, 1008, 1010, 1017, 1022, 1023, 1024, 1025, 1030, 1034, 1038, 1039, 1054
2786	25252968 (13787, 13788)	Novel Protein sim. GBank gi 3322607 (AE001213) - serine hydroxymethyltransferase (glyA) [Treponema pallidum]	Contains protein domain (PF00464) - Serine hydroxymethyltransferase	UNCLASSIFIED	1026
2787	369911774 (20433, 20434)	Novel Protein sim. GBank gi 132245 sp P24718 RECF_ACTPL - RECF PROTEIN	Contains protein domain (PF00470) - RecF protein	UNCLASSIFIED	1016
2788	80047309 (7343, 7344)	Novel Protein sim. GBank gi 730753 sp P38899 SMC2_YEAST - CHROMOSOME SEGREGATION PROTEIN SMC2 (DA-BOX PROTEIN SMC2)	Contains protein domain (PF00470) - RecF protein	UNCLASSIFIED	1003, 1004, 1008, 1010, 1022, 1023, 1027, 1039, 1044

2789 17336)	23332078 (17335, Novel Protein sim. GBank gi 4490317 emb CAB38808.1 - (AL035678) putative protein [Arabidopsis thaliana]	Contains protein domain (PF00481) - Protein phosphatase 2C	UNCLASSIFIED	1053
2790 5118)	10847409 (5117, Novel Protein sim. GBank gi 1170055 sp P41441 GSPF_ECOLI - PUTATIVE GENERAL SECRETION PATHWAY PROTEIN F (PROTEIN TRANSPORT PROTEIN HOFFY)	Contains protein domain (PF00482) - Bacterial type II secretion system protein F domain	UNCLASSIFIED	1017
2791 11492)	65458979 (11491, Novel Protein sim. GBank gi 130213 sp P22609 PILC_PSEAE - FIMBRIAL ASSEMBLY PROTEIN PILC	Contains protein domain (PF00482) - Bacterial type II secretion system protein F domain	UNCLASSIFIED	1054
2792 9252)	71458849 (9251, Novel Protein sim. GBank gi 3546660 emb CAA20571.1 - (AL031394) carbonate dehydratase - like protein [Arabidopsis thaliana]	Contains protein domain (PF00484) - Prokaryotic-type carbonic anhydrases	UNCLASSIFIED	1008, 1026
2793 18580)	56512548 (18579, Novel Protein sim. GBank gi 91984 pir A26702 - arginase (EC 3.5.3.1), hepatic - rat	Contains protein domain (PF00481) - Arginase family	UNCLASSIFIED	1010, 1049
2794 12916)	29684760 (12915, Novel Protein sim. GBank gi 1169692 sp P41813 FKH2_YEAST - FORK HEAD PROTEIN HOMOLOG 2	Contains protein domain (PF00498) - Forkhead- associated (FHA) domain	UNCLASSIFIED	1006, 1009, 1022
2795 8264)	78374720 (8263, Novel Protein sim. GBank gi 5102795 emb CAB45210.11 - (AL079308) hypothetical protein [Streptomyces coelicolor].	Contains protein domain (PF00498) - Forkhead- associated (FHA) domain	UNCLASSIFIED	1008
2796 8912)	27989751 (8911, Novel Protein sim. GBank gi 430947 (U01103) - PSI type III chlorophyll a/b-binding protein [Arabidopsis thaliana]	Contains protein domain (PF00504) Chlorophyll A-B binding proteins	UNCLASSIFIED	1003
2797 2138)	80210898 (2137, Novel Protein sim. GBank gi 1176229 sp P42810 YHE3_PSEAE - HYPOTHETICAL 64.8 KD PROTEIN IN HEMM- HEMA INTERGENIC REGION (ORF3)	Contains protein domain (PF00515) - TPR Domain	UNCLASSIFIED	1004, 1039
2798 6432)	80503185 (6431, Novel Protein sim. GBank gi 134515 sp P17883 SKI3_YEAST - SUPERKILLER 3 PROTEIN	Contains protein domain (PF00515) - TPR Domain	UNCLASSIFIED	1004, 1012, 1022, 1038

				UNCLASSIFIED	1024
2799	20474337 (717, 718)	Novel Protein sim. GBank gil2117303 emb CAE09127.1 - (Z95620) n-acetylglucosaminyl-phosphatidylinositol [Schizosaccharomyces pombe]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	
2800	87467604 (22765, 22766)	Novel Protein sim. GBank gil2911108 emb CAA17492 - (AL021957) hypothetical protein Rv2188c [Mycobacterium tuberculosis]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	1018
2801	13084838 (14747, 14748)	Novel Protein sim. GBank gil3063872 emb CAA18552 - (AL022486) hypothetical protein MLCB1883.03 [Mycobacterium leprae]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	1027
2802	17898561 (22239, 22240)	Novel Protein sim. GBank gil3249551 (AF010182) - glycosyltransferase WbpY [Pseudomonas aeruginosa]	Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	1017
2803	66694692 (18385, 18386)		Contains protein domain (PF00534) - Glycosyl transferases group 1	UNCLASSIFIED	1009
2804	78386436 (6447, 6448)	Novel Protein sim. GBank gil2494771 sp Q53044 GLNB_RHORU - NITROGEN REGULATORY PROTEIN P-II	Contains protein domain (PF00534) - Nitrogen regulatory protein P-II	UNCLASSIFIED	1026
2805	11076638 (22251, 22252)	Novel Protein sim. GBank gil4512117 gb AAD21622.1 - (AF128399) succinyl CoA synthetase beta subunit [Pseudomonas aeruginosa]	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	1024
2806	23331990 (18593, 18594)	Novel Protein sim. GBank gil5304835 emb CAE46076.1 - (AJ243439) ATP citrate lyase [Cyanophora paradoxa]	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	1053
2807	79854118 (19147, 19148)	Novel Protein sim. GBank gil13315 sp P25464 ACVS_CEPAC - DELTA-(L-ALPHA-AMINOADIPYL)-L-CYSTEINYLD-VALINE SYNTHETASE (ACV SYNTHETASE) (ACVS)	Contains protein domain (PF00550) - Phosphopantetheine attachment site	UNCLASSIFIED	1044
2808	78808519 (19269, 19270)	Novel Protein sim. GBank gil549741 sp P36047 SD22_YEAST - PROTEIN PHOSPHATASES PP1 REGULATOR SUBUNIT SDS22	Contains protein domain (PF00560) - Leucine Rich Repeat	UNCLASSIFIED	1009, 1022, 1044

2809	20286931 (13873) 13874)	Novel Protein sim. GBank gi 35581852 emb CAA208081 - (AL031541) putative rRNA methylase [Streptomyces coelicolor]	Contains protein domain (PF00588) - Spoo rRNA Methylase family	UNCLASSIFIED	1029, 1034
2810	80248618 (14857, 14858)	Novel Protein sim. GBank gi 173284 isp Q10815 YX25 MYCTU - PROBABLE INTEGRASE/RECOMBINASE CY274.25C	Contains protein domain (PF00569) - "Phage" integrase family	UNCLASSIFIED	1003, 1006, 1008, 1009, 1016, 1026, 1049
2811	28382058 (7061, 7062)	Novel Protein sim. GBank gi 1705505 isp P54729 BS4_MOUSE - BS4 PROTEIN	Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	1010, 1044
2812	78749547 (5011, 5012)	Novel Protein sim. GBank gi 2132222 pir IS65193 - hypothetical protein YPL181w - yeast (Saccharomyces cerevisiae)	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	1022, 1044
2813	19752799 (6467, 6468)	Novel Protein sim. GBank gi 4584215 emb CAB40628.1 - (Y14999) Bip2 protein [Drosophila melanogaster]	Contains protein domain (PF00628) - PHD-finger	UNCLASSIFIED	1038
2814	86471760 (5377, 5378)	Novel Protein sim. GBank gi 4585584 emb CAB40852.1 - (AL049628) putative DNA-binding protein [Streptomyces coelicolor]	Contains protein domain (PF00633) - Helix-hairpin- helix motif.	UNCLASSIFIED	1004, 1025, 1026, 1029
2815	57306232 (10139, 10140)		Contains protein domain (PF00672) - Domain found in bacterial signal proteins	UNCLASSIFIED	1017, 1022, 1026
2816	20287844 (7129, 7130)	Novel Protein sim. GBank gi 3281719 emb CAA06651 - (Z85982) infC [Mycobacterium tuberculosis]	Contains protein domain (PF00707) - Translation initiation factor IF-3	UNCLASSIFIED	1034
2817	78386071 (2879, 2880)	Novel Protein sim. GBank gi 1001829 dbj BAA10710 - (D64005) hypothetical protein [Synechocystis sp.]	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	1026
2818	78086204 (1857, 1858)	Novel Protein sim. GBank gi 2496818 isp Q10940 YvS4_CAEEL - HYPOTHETICAL 13.1 KD PROTEIN B0310.4 IN CHROMOSOME X	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	1033, 1055
2819	87890634 (20497, 20498)	Novel Protein sim. GBank gi 897612 (L41589) - epithelial mucin [Hylobates lar]	Contains protein domain (PF00711) - Beta defensins	UNCLASSIFIED	1015, 1018, 1019, 1024, 1048

2820	30243803 (1173, 1174)	Novel Protein sim. GBank gi 3342384 (AF063001) - B-N-acetylhexosaminidase [Streptomyces plicatus]	Contains protein domain (PF00728) - Glycosyl hydrolase family 20	UNCLASSIFIED	1026
2821	78730724 (8449, 8450)	Novel Protein sim. GBank gi 3288113 emb CA809803 - (Z97067) elongation factor 1-beta [Beta vulgaris]	Contains protein domain (PF00736) - EF-1 guanine nucleotide exchange domain	UNCLASSIFIED	1008
2822	27981234 (5719, 5720)		Contains protein domain (PF00746) - Gram positive anchor	UNCLASSIFIED	1006
2823	27977901 (8977, 8978)		Contains protein domain (PF00746) - Gram positive anchor	UNCLASSIFIED	1006
2824	56648176 (18835, 18836)	Novel Protein sim. GBank gi 603957 dbj BAAC07895 - (D43951) KIAA0099 is related to D.melanogaster pumilio gene. [Homo sapiens]	Contains protein domain (PF000806) - Pumilio-family RNA binding domains (aka PUM-HD, Pumilio homology domain)	UNCLASSIFIED	1024
2825	80052477 (4741, 4742)	Novel Protein sim. GBank gi 732333 sp P39861 YWCH_BAGCU - HYPOTHETICAL 36.6 KD PROTEIN IN QOXD-VPR INTERGENIC REGION	Contains protein domain (PF00818) - Ice nucleation protein repeat	UNCLASSIFIED	1010, 1024, 1026, 1030, 1039
2826	20286961 (17157, 17158)	Novel Protein sim. GBank gi 1402520 dbj BAA11042 - (D67027) transposase of IS1164 [Rhodococcus rhodochrous]	Contains protein domain (PF00872) - Transposase, Mutator family	UNCLASSIFIED	1034
2827	79615760 (14685, 14686)	Novel Protein sim. GBank gi 1480789 (U62766) - transposase [Mycobacterium branderii]	Contains protein domain (PF00872) - Transposase, Mutator family	UNCLASSIFIED	1003, 1022
2828	79184257 (3975, 3976)	Novel Protein sim. GBank gi 2078069 emb CA808502 - (Z95210) hypothetical protein Rv0920c [Mycobacterium tuberculosis]	Contains protein domain (PF00872) - Transposase, Mutator family	UNCLASSIFIED	1027
2829	21632022 (15177, 15178)	Novel Protein sim. GBank gi 549115 sp P35677 TRA_CORDI - PROBABLE TRANSPONASE FOR INSERTION SEQUENCE ELEMENT	Contains protein domain (PF00872) - Transposase, Mutator family	UNCLASSIFIED	1023, 1024

2830	79175532 (9425, 9426)	Novel Protein sim. GBank gi 128806 sp P25197 NOLG_RHIME - NODULATION PROTEIN NOLG PRECURSOR	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1031
2831	39550301 (2457, 2458)	Novel Protein sim. GBank gi 1655901 dbj BAA18811 - (D90917) acriflavine resistance protein [Synechocystis sp.]	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1039
2832	25238645 (4717, 4718)	Novel Protein sim. GBank gi 1736782 dbj BAA15930 - (D90845) Acriflavin resistance protein F (EnvD protein). [Escherichia coli]	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1003
2833	80422500 (6191, 6192)	Novel Protein sim. GBank gi 1736782 dbj BAA15930 - (D90845) Acriflavin resistance protein F (EnvD protein). [Escherichia coli]	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1025
2834	94679072 (3621, 3622)	Novel Protein sim. GBank CeoB [Burkholderia cepacia]	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1022, 1031
2835	95295877 (2513, 2514)	Novel Protein sim. GBank gi 2506614 sp P38054 YBDE_ECOLI - HYPOTHETICAL 114.7 KD PROTEIN IN NFRB-PHEP INTERGENIC REGION	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1044
2836	11090814 (14255, 14256)	Novel Protein sim. GBank gi 2506614 sp P38054 YBDE_ECOLI - HYPOTHETICAL 114.7 KD PROTEIN IN NFRB-PHEP INTERGENIC REGION	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1004
2837	66051809 (11295, 11296)	Novel Protein sim. GBank gi 3025164 sp P76398 YEGN_ECOLI - HYPOTHETICAL 112.1 KD PROTEIN IN ALKA-BAES INTERGENIC REGION	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1010
2838	78680377 (15003, 15004)	Novel Protein sim. GBank gi 3025164 sp P76398 YEGN_ECOLI - HYPOTHETICAL 112.1 KD PROTEIN IN ALKA-BAES INTERGENIC REGION	Contains protein domain (PF000873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1003

2839	85818244 (1755, 1756)	Novel Protein sim. GBank gil4206630 (AF067954) - putative inner membrane proton/cation antipporter SIIA [Salmonella typhimurium]	Contains protein domain (PF00873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1019
2840	11093010 (17941, 17942)	Novel Protein sim. GBank gil4206630 (AF067954) - putative inner membrane proton/cation antipporter SIIA [Salmonella typhimurium]	Contains protein domain (PF00873) - AcrB/AcrD/AcrF family	UNCLASSIFIED	1006
2841	80026023 (15249, 15250)	Novel Protein sim. GBank gil1341801sp P15401 SACY_BACSU - LEVANSUCRASE AND SUCRASE SYNTHESIS OPERON ANTITERMINATOR	Contains protein domain (PF00874) - Transcriptional antiterminator bgfG family	UNCLASSIFIED	1004, 1006, 1008, 1010, 1024, 1026, 1034, 1044, 1053, 1054
2842	78788983 (21055, 21056)	Novel Protein sim. GBank gil1723641sp P53202 YG13_YEAST - CULLIN B	Contains protein domain (PF00888) - Culin family	UNCLASSIFIED	1003, 1009, 1026
2843	95358293 (8615, 8616)	Novel Protein sim. GBank gil25015341sp P75802 YLIG_ECOLI - HYPOTHETICAL 49.6 KD PROTEIN IN MOEA-DACC INTERGENIC REGION	Contains protein domain (PF00919) - Uncharacterized protein family UPF0004	UNCLASSIFIED	1058
2844	78925859 (15175, 15176)	Novel Protein sim. GBank gil25015341sp P75802 YLIG_ECOLI - HYPOTHETICAL 49.6 KD PROTEIN IN MOEA-DACC INTERGENIC REGION	Contains protein domain (PF00919) - Uncharacterized protein family UPF0004	UNCLASSIFIED	1039
2845	27978032 (12357, 12358)	Novel Protein sim. GBank gil1386991sp P09963 LYCV_BPP22 - LYSOZYME (LYSIS PROTEIN)	Contains protein domain (PF00959) - Lysozyme	UNCLASSIFIED	1022
2846	84452949 (2535, 2536)	Novel Protein sim. GBank gil347493 (L22072) - PET112 protein [Saccharomyces cerevisiae]	Contains protein domain (PF01162) - PET112 family	UNCLASSIFIED	1003, 1004, 1006, 1008, 1009, 1010, 1012, 1022, 1023, 1024, 1025, 1034, 1038, 1039, 1044, 1049

2847	78767224 (6677, 6678)	Novel Protein sim. GBank gi 1169716 sp P44788 SUN_HAEIN - SUN PROTEIN (FMU PROTEIN)	Contains protein domain (PF01189) - NOL1/NOP2/sun family	UNCLASSIFIED	1003, 1008, 1009, 1022, 1026
2848	11076699 (10097, 10098)	Novel Protein sim. GBank gi 55457 (L26406) - orf1 [Methyllobacterium extorquens]	Contains protein domain (PF01381) - Helix-turn-helix	UNCLASSIFIED	1010
2849	78969046 (9431, 9432)	Novel Protein sim. GBank gi 1033263 pir A53330 - collagen alpha 2(IX) chain precursor - mouse	Contains protein domain (PF01391) - Collagen triple helix repeat (20 copies)	UNCLASSIFIED	1030, 1038
2850	71761016 (19559, 19560)	Novel Protein sim. GBank gi 283032 pir S22456 - hydroxyproline-rich glycoprotein - perennial teosinte	Contains protein domain (PF01402) - Helix-turn-helix protein, copG family	UNCLASSIFIED	1026
2851	39727058 (13385, 13386)		Contains protein domain (PF01405) - Photosystem II reaction centre T protein	UNCLASSIFIED	1026
2852	86471839 (15641, 15642)	Novel Protein sim. GBank gi 174501 sp P43816 SYC_HAEIN - CYSTEINYL TRNA SYNTHETASE (CYSTEINE-TRNA LIGASE) (CYRSRS)	Contains protein domain (PF01406) - tRNA synthetases class I (C)	UNCLASSIFIED	1029
2853	8488670 (11865, 11866)	Novel Protein sim. GBank gi 3122887 sp P98662 SYC_MYCTU - CYSTEINYL-TRNA SYNTHETASE (CYSTEINE-TRNA LIGASE) (CYRSRS)	Contains protein domain (PF01406) - tRNA synthetases class I (C)	UNCLASSIFIED	1024
2854	86471572 (1873, 1874)	Novel Protein sim. GBank gi 183601 emb CAB06643 - (Z85982) phes - [Mycobacterium tuberculosis]	Contains protein domain (PF01409) - tRNA synthetases class II (F)	UNCLASSIFIED	1019
2855	86380243 (2495, 2496)		Contains protein domain (PF01409) - tRNA synthetases class II (F)	UNCLASSIFIED	1053
2856	86473215 (3793, 3794)	Novel Protein sim. GBank gi 3122879 sp O07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)	Contains protein domain (PF01411) - tRNA synthetases class II (A)	UNCLASSIFIED	1029
2857	28383962 (5933, 5934)	Novel Protein sim. GBank gi 3122879 sp O07438 SYA_MYCTU - ALANYL-TRNA SYNTHETASE (ALANINE-TRNA LIGASE) (ALARS)	Contains protein domain (PF01411) - tRNA synthetases class II (A)	UNCLASSIFIED	1008, 1016, 1022